随 机 过 程

方兆本 缪柏其 编著

中国科学技术大学出版社
1992・合肥

（皖）新登字08号

随　机　过　程

方兆本　缪柏其　编著

中国科学技术大学出版社出版

（安徽省合肥市金寨路96号，邮政编码：230026）

金寨县印刷厂印刷

安徽省新华书店发行

开本：850×1168/32　　印张：5.625　　字数：143千

1993年1月第1版　　1996年3月第2次印刷

印数：4001—7000册

ISBN7-312-00379-6/O·122　　定价：5.50元

内 容 提 要

本书介绍在应用中经常遇到的几种基本随机过程，如泊松过程，更新过程，马尔柯夫过程，平稳过程，布朗运动，以及线性随机微分方程，材料丰富．每章结合大量有实际背景的例子来解释基本概念，并配有一定量的习题．

本书可作为理、工科大学生和研究生的教学用书或教学参考书，也可作为工程技术人员应用随机过程的入门参考书．

前　　言

本书是作者与中国科学技术大学数学系同事们多年从事本课程教学的积累。它是为大学理工科本科生、研究生概率统计公共课所编写的教材、是陈希孺教授所著的《概率论与数理统计》的续篇。本书可以作非数学系应用数学辅修专业的随机过程课的教本，主要讨论随机过程的基本理论及其应用。每章正文之后有配套的习题供读者练习。随机过程是一门应用性很强的学科、各个领域中的科技工作者都能从中发现有启发性的模型。单纯照搬模型比较容易，难的是在实际问题中简化条件提炼出恰当的随机模型。学习这门课程应在这方面多做努力，这也是本书编写的宗旨。

学习本书要有微积分和初等概率论的基础，兼顾内容阐述的需要和学生的实际接受能力书中用到了矩母函数、生成函数、复变函数、微分方程求解及矩阵代数等数学工具。对某些工具不熟悉的读者可以跳过证明推理直接阅读有关结论。这并不影响对本课程的基本了解。

使用本教材的教员自然可以根据课时限制及各系各科的不同需求而有所侧重。比如，删去 2.3 节，3.3 节中定理 3.3 和 3.4 的证明，3.4、3.6 节、4.2 节的定理 4.2、4.4 节、以及 5.3 节将仍是一份应用随机过程的 ABC 教材。书末曾备有习题答案或提示，但考虑到附在书后出版对教与学无益故予以删除。作者感谢胡太忠同志帮助演算了书中的全部习题。

作　者

1991 年 9 月于合肥

目　　录

第一章 引 论

1.1 引 言

1.1.1 基本概念和例子

随机过程是对一连串随机事件间动态关系的定量描述。它是在自然科学、工程科学、社会科学各领域研究随机现象的有力工具。其应用包罗万象：气象预报、天文观测、通讯工程、原子物理、宇航遥控、生物医学、管理科学、运筹决策、计算机科学、经济分析、人口理论、可靠性与质量控制等许许多多领域都已离不开用随机过程的理论来建立各种数学模型。

一般，把一族随机变量定义为随机过程。英文叫 stochastic process。"stochastic" 一词源于希腊语 "στοχαστικοξ" 意思是"猜"。但这门科学不是乱猜。在研究随机过程时人们透过表面的偶然性找出必然的内在规律并以概率的形式来描述这些规律。从偶然中悟出必然正是这一学科的魅力所在。

随机过程的早期历史属于物理领域。人们可以追溯到 Gibbs，Boltzman，Poincaré 等人在统计力学中的研究以及后来 Einstein，Wiener，Lévy 等人对布朗运动的开创性工作。而 Erlang 等则在电话流中研究了 Poisson 过程。而整个学科的理论基础则是由 Kolmogorov 和 Doob 奠定的。"stochastic" 这一用词也在这时流行。生灭过程是 Feller 首先引进的。Cramer 和 Lévy 研究了平稳过程。Xinchin，Palm 发展了排队论中的过程理论。Doob 则研究 Markov 过程和鞅。这些都是早期研究

的重要里程碑。目前，这一学科仍在理论和应用两方面以空前的深度和广度在迅速发展着。下面对随机过程作正式定义：

定义 1.1 随机过程就是一族随机变量$\{X(t),\ t\in T\}$，其中t是参数它属于某个指标集T，T称为是参数集。

一般，t代表时间。当$T=\{0,1,2,\cdots\}$时也称随机过程为随机序列。对$X(t)$可以这样看：随机变量是定义在空间Ω上的，所以是随t与$\omega\in\Omega$而变化的，于是可以记为$X(t,\omega)$。当固定一次随机试验，即取定$\omega_0\in\Omega$时，$X(t,\omega_0)$就是一条样本路径它是t的函数。它可能是连续的，也可能是有间断点和跳跃的。这是我们通常所观察到的过程。另一方面固定了时间$t=t_0$，$X(t_0,\omega)$就是一个随机变量，其取值随着随机试验的结果而变化。变化有一定的规律、叫做概率分布。随机过程$X(t)$取的值称作是过程所处的状态。状态的全体称为状态空间。根据T及状态空间的不同我们可以对过程进行分类。依照状态空间可分为连续状态和离散状态；依参数集T，当T为有限集或可列集则称之为离散参数过程否则称为是连续参数过程。当T是高维向量则$X(t)$称作是随机场。

例 1.1 英国植物学家 Brown 注意到漂浮在液面上的微小粒子不断进行无规则的运动。这种运动叫做 Brown 运动。它是分子大量随机碰撞的结果。若记$(X(t),Y(t))$为粒子在平面坐标上的位置，则它是平面上的 Brown 运动。在统计物理中对它有深入的研究。

例 1.2 一醉汉在路上行走，以概率p前进一步，概率$1-p$后退一步。以$X(t)$记他在街上的位置则$X(t)$就是直线上的随机游动。

例 1.3 神经细胞在细胞膜的位势达到某一临界值C时就要兴奋。刺激和抑制两种脉冲以一定的速率(比如 Poisson 过程)抵达细胞。前者使位势升高；后者使位势降低。升降的幅度服从相同的分布$H(x)$。神经细胞在兴奋过后位势恢复到 0，过程再度重

复．记 T_i 为两次兴奋的间隔时间，并记 $X(t)$ 为时刻 t 时细胞膜的位势则过程的一次实现如图 1.1 所示：



图 1.1 神经细胞膜的位势

例 1.4 到达总机交换台的呼叫次数为 Poisson 过程。每次呼叫是相互独立的，而间隔时间服从指数分布。交换台在同一时刻只能接通 K 个呼叫。人们常要了解在某一时刻的排队长度以及呼叫的平均等待时间。这是一种排队模型。

例 1.5 流行病学的研究中有如下模型：在时刻 0 时易感人群大小为 $X(0)$，$Y(0)$ 是已受传染的人数。假定易感人群被传染的概率为 p，则经过一段传染周期后（记之为单位时间）$X(0)$ 中有 $X(1)$ 没有染上病而 $Y(1)$ 却受到传染。传染过程一直漫延到再没有人会染上这种流行病时停止。于是 $X(t)=X(t+1)+Y(t+1)$，且当 $j\leqslant i$ 时有

$$P(X(t+1)=j|X(t)=i)=\binom{i}{i-j}p^{i-j}(1-p)^{j}.$$

$\{X(t),\ t=0,1,2,\cdots\}$ 就是以上式为状态转移概率的 Markov 过程。

例 1.6 设 $X(t)$ 为信号流，它满足方程

$$X(t+1)=aX(t)+\varepsilon(t),$$

其中 a 为实参数，$\varepsilon(t)$ 代表误差。真正的信号 $X(t)$ 并不可能观察到。人们所能观察到的是

$$Y(t)=X(t)+Z(t),$$

$Z(t)$为噪声。从观察值$Y(t)$出发，检测出$X(t)$是通讯工程中的重要课题。这是过程的预测与滤波的问题。

例 1.7　水库库容调度。记$Y(t)$为$(t, t+1)$年间的水库蓄水量，它是随机的。M为每年年底固定的清库泄洪量，$X(t)$为泄洪之后的库容量。则$X(t+1)=\min\{X(t)+Y(t), K\}-\min\{X(t)+Y(t), M\}$其中$K$为大坝的设计库容。过程$Y(t)$为一Markov过程。把水库库容改为商店仓库的库存模型类似。

例 1.8　记$X(t)$为时刻t的商品价格。若$X(t)$适合线性模型

$$\begin{aligned}&X(t)+\alpha_1X(t-1)+\alpha_2X(t-2)+\cdots\alpha_pX(t-p)\\&\quad=Z(t)+\beta_1Z(t-1)+\cdots+\beta_qZ(t-q).\end{aligned}$$

其中α_i，β_j为实参数，$Z(t)$为独立同分布的不可观察的随机变量。则$X(t)$服从所谓ARMA模型——混合自回归滑动平均模型。这是在经济预测中十分有用的时间序列模型。

1.1.2　有限维分布和数字特征

知道了分布函数$F_X(x)=P(X\leqslant x)$就能了解随机变量X。类似的，对随机过程$\{X(t), t\in T\}$定义$F_t(x)=P\{X(t)\leqslant x\}$为过程的一维分布，$X(t)$的期望$E[X(t)]$为过程的均值函数记作$\mu_X(t)$。而$\mathrm{Var}[X(t)]$则被定义为过程的方差函数。不仅如此，我们还需要了解随机变量$X(t_1)$与$X(t_2)$的联合分布$P\{X(t_1)\leqslant x_1, X(t_2)\leqslant x_2\}$，这也就是过程在$t_1$，$t_2$两不同时刻值的联合二维分布记作$F_{t_1, t_2}(x_1, x_2)$。$\mathrm{Cov}(X(t_1), X(t_2))$则称为是过程的协方差函数，记作$r_X(t_1, t_2)$。于是有

$$r_X(t_1,t_2)=E\{(X(t_1)-\mu_X(t_1))(X(t_2)-\mu_X(t_2))\}. \tag{1.1}$$

协方差函数显然有对称性，即对任何s, t有$r_X(t,s)=r_X(s,t)$。当$t_1=t_2=t$时。$r_X(t,t)=\mathrm{Var}X(t)$。容易证明协方差函数是非负定的，也就是说对任何$t_1, t_2, \cdots, t_n\in T$及任意实数$b_1, \cdots, b_n$我们恒有

$$\sum_{i=1}^{n}\sum_{j=1}^{n}b_i b_j r_X(t_i,t_j)\geqslant 0. \tag{1.2}$$

因为协方差运算有线性性质，由

$$\mathrm{Var}\left(\sum_{j=1}^{n} b_j X(t_j)\right)=\mathrm{Cov}\left(\sum_{i=1}^{n} b_i X(t_i),\sum_{j=1}^{n} b_j X(t_j)\right)$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{n}b_i b_j r_X(t_i,t_j)$$

知(1.2)式成立故它是非负定的。进而，我们定义随机过程的有限维分布族它是 $F_{t_1,t_2,\cdots t_n}(x_1,\cdots,x_n)$，$t_1,\cdots,t_n\in T$ 全体。其中

$$F_{t_1,\cdots,t_n}(x_1,\cdots,x_n)=P\{X(t_1)\leqslant x_1,\cdots,X(t_n)\leqslant x_n\}. \tag{1.3}$$

知道了随机过程的有限维分布族就知道了过程$\{X(t),\ t\in T\}$中任意 n 个随机变量的联合分布。也就完全了解了这些变量之间的相互依赖关系。有限维分布也有对称性，它与变量 $X(t_1),\cdots,X(t_n)$ 的排列次序无关。对$\{1,\cdots,n\}$的任一置换$(i_1,\cdots,i_n)$有

$$F_{t_{i_1},\cdots,t_{i_n}}(x_{i_1},\cdots,x_{i_n})=F_{t_1,\cdots,t_n}(x_1,\cdots,x_n).$$

有限维分布还有相容性，意思是当某些 $x_i\to\infty$时高维分布的边缘分布与相应的低维分布是一致的。也即对 $m<n$ 有

$$F_{t_1,\cdots,t_m,t_{m+1},\cdots,t_n}(x_1,\cdots,x_m,\infty,\cdots,\infty)$$

$$=F_{t_1,\cdots,t_m}(x_1,\cdots,x_m).$$

有趣的是若一族给定的分布函数有上述对称性和相容性则保证了存在一个随机过程$\{X(t),t\in T\}$使它的有限维分布族正好就是给定的分布函数族。这是 1931 年由 Kolmogorov 证明的相容性基本定理。

例 1.9 设 X_n 为第 n 次独立地扔一六面骰子的结果。则$\{X_n,\ n\geqslant 1\}$为一随机过程，参数集 T 为$\{1,2,3,\cdots\}$。而状态空间为$\{1,2,3,4,5,6\}$。序列 3，1，4，6，3，2，5，5，1，…就是$\{X_n,\ n\geqslant 1\}$的一次可能的实现。均值函数 $EX_n=EX_1\equiv 3.5$，除方差函数为 17.5 外协方差函数恒等于 0。任何有限维分布 $F_{n_1,\cdots,n_k}(x_1,\cdots,x_k)=F(x_1)\cdots F(x_k)$其中 $F(x)$为 X_1 的分布函数。

当然这种独立情形是太过于简单了。对于大多数随机过程，

$X(t)$之间是相依的。研究他们之间的相依关系是我们的主要课题。根据相依关系的不同，人们可以研究随机过程的不同类型。

1.1.3 平稳过程和独立增量过程

如果两个随机变量X_1, X_2的分布函数$F_{X_1}(x)$与$F_{X_2}(x)$对任何x都是相等的则称它们是等分布的记作$X_1 \stackrel{d}{=} X_2$。类似的如果一个随机向量$\boldsymbol{X}=(X_1,\cdots,X_n)$与另一随机向量$\boldsymbol{Y}=(Y_1,\cdots,Y_n)$有相同的联合分布则也称他们是等分布的记作$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{Y}$。有一类重要的随机过程，它处于某种概率平衡状态其主要性质只与变量$X(t)$之间的时间间隔有关而与我们考查的起始点无关。这类过程叫做平稳过程。具体而言我们有

定义 1.2 如果随机过程$X(t)$对任意的$t_1,\cdots,t_n \in T$和任何h有

$$(X(t_1+h),\cdots,X(t_n+h)) \stackrel{d}{=} (X(t_1),\cdots,X(t_n)), \quad (1.4)$$

则称作是严格平稳的。

条件(1.4)很强也不易验证。所以退而求其次有所谓宽平稳或二阶平稳过程。引入

定义 1.3 如果随机过程$X(t)$的所有二阶矩存在并有$EX(t)=m$及协方差函数$r_X(t,s)$只与时间差$t-s$有关则称作是宽平稳的或二阶矩平稳的。

对于宽平稳过程由于对$-\infty<s,t<\infty, r_X(s,t)=r_X(0,t-s)$所以可以记之为$R_X(t-s)$。显然对所有$t$ $R_X(-t)=R_X(t)$，即为偶函数。所以$R_X(t)$的图形是关于坐标轴对称的。其在0点的值$R_X(0)$就是过程的方差函数，即$R_X(0)=\mathrm{Var}X(t)$。

尽管$X(t)$之间常常不是相互独立的，但人们可以假定过程的增量之间是相互独立的。这就是

定义 1.4 如果对任何$t_1,\cdots,t_n \in T$, $t_1<t_2<\cdots<t_n$随机变量$X(t_2)-X(t_1), X(t_3)-X(t_2),\cdots,X(t_n)-X(t_{n-1})$是相互独立的

则$X(t)$称为是独立增量过程。如果进一步有对任何t_1，t_2 $X(t_1+h)-X(t_1)\stackrel{d}{=}X(t_2+h)-X(t_2)$则过程称为是有平稳独立增量的过程。

可以证明平稳独立增量过程的均值函数一定是t的线性函数。我们以后要介绍的Poisson过程和Brown运动都是这类过程。这两类过程是过程理论中的两块最重要的基石。

例 1.10 设Z_i，$i=0,1,2,\cdots$是一串独立同分布的随机变量定义$X_n=\sum_{i=0}^{n}Z_i$则过程$\{X_n, n\geqslant 0\}$就是独立增量过程。一般称X_n为独立和。

重要的相依类型还有Markov链及Markov过程(这构成了第三章的主题)、更新过程和鞅等。更新过程是运筹学、排队论等管理学科中的重要工具；鞅是近代概率论及随机过程理论中的重要概念。受篇幅及课时限制都无法一一展开讨论。读者应注意到过程的分类不是绝对的，一个具体的过程可以同时属于上述多种类型。比如第二章要讨论的Poisson过程既有独立增量又有平稳增量。它既是连续时间的Markov链又是一类特殊的更新过程。若过程的参数为λ则Poisson过程$X(t)$减去均值函数即$X(t)-\lambda t$还是一个鞅。

1.2 条件期望、矩母函数

1.2.1 条件期望

要研究随机过程离不开讨论一族随机变量相互之间的关系，这时常要用到条件概率和条件期望这些基本的概念和工具。本节将对有关概念和工具作一简要回顾。学过概率论的读者都一定知道什么是事件的条件概率，即给定事件B发生时事件A发生的条件概率$P(A|B)$等于$P(A\cap B)/P(B)$。但究竟给了随机变量

Y 的取值后另一随机变量 X 的条件期望是什么却未必能说清楚。先看离散型随机变量 X 和 Y：一般对所有使 $P\{Y=y\}>0$ 的 y 定义给定 $Y=y$ 时 X 取 x 的条件概率为

$$P\{X=x|Y=y\}=\frac{P\{X=x,Y=y\}}{P\{Y=y\}}. \tag{1.5}$$

而给定 $Y=y$，X 的条件分布函数则定义为

$$F(x|y)=P\{X\leqslant x|Y=y\}. \tag{1.6}$$

相应的，给定 $Y=y$，X 的条件期望定义为

$$E(X|Y=y)=\sum_x xP\{X=x|Y=y\}. \tag{1.7}$$

对一般的连续型随机变量 $P(Y=y)$ 往往为 0，比式(1.5) 没有意义我们怎么来定义条件期望呢？常用的办法是如果对任何包含 y 的小区间总有 $P(Y\in\Delta y)=0$，则定义 $P(X\in A|Y=y)=0$。否则，若 $P(Y\in\Delta y)>0$ 就可以定义

$$P(X\in A|Y=y)=\lim_{\Delta y\downarrow 0}P(X\in A|Y\in\Delta y).$$

这里 $\Delta y\downarrow 0$ 的意思是使包含 y 的小区间的长度缩小为 0。除了个别例外的 y 值这一极限总是存在。相应地，定义给定 $Y=y$ 时，X 的条件分布函数为

$$P(X\leqslant x|Y=y)=\lim_{\Delta y\downarrow 0}P(X\leqslant x|Y\in\Delta y), \tag{1.8}$$

记作 $F(a|y)$。进一步，如果存在一非负函数（记为）$f(x|y)$ 使得对任何集合 A 恒有

$$P(X\in A|Y=y)=\int_A f(x|y)\mathrm{d}x, \tag{1.9}$$

且 $\int f(x|y)\mathrm{d}x=1$ 则 $f(x|y)$ 称为是给定 $Y=y$ 时 X 的条件密度。不难看出条件分布函数是关于条件密度对变量 x 从 $-\infty$ 到 x 的积分，即

$$F(x|y)=\int_{-\infty}^{x}f(x|y)\mathrm{d}x. \tag{1.10}$$

如果把 X 与 Y 的联合密度 $f(x,y)$ 看作质量为 1 的平板的密度

(面密度)则条件密度 $f(x|y)$ 就是固定 $Y=y$ 时的线密度。它们之间的关系是

$$f(x,y)=f(x|y)f(y), \tag{1.11}$$

其中 $f(y)$为随机变量 Y 的边缘密度。此时，给定 $Y=y$，X 的条件期望为

$$E(X|Y=y)=\int xf(x|y)\mathrm{d}x. \tag{1.12}$$

(1.7)式与(1.12)式本质上是一回事，在更深一点的课程中人们把它们统一记为

$$E(X|Y=y)=\int x\mathrm{d}F(x|y). \tag{1.13}$$

在行文不发生混淆时，也常记为 $E(X|Y)$。严格地讲(1.13)表示一个数值(与 y 有关)而 $E(X|Y)$则为随机变量。

例 1.11 扔一硬币出现正面的概率为 p，独立地作投币试验。记 S 为 n 次试验中出现正面的总次数，并设首次出现正面是在第 T 次试验。问题是求给定 n 次试验中仅出现了一次正面时随机变量 T 的条件概率分布，也即 $P(T=k|S=1)$。

一种想法认为给定了正面出现一次它会以相同的机会在 n 次试验中的任一次出现所以 T 的条件概率分布应该是 $1,2,\cdots,n$ 上的均匀分布。但有人也许会争辩：因为给定了只出现一次正面它很可能会较早出现。让我们来实际算一算：事件 $\{T=k,S=1\}$只包含一个序列，它仅在第 k 次试验时结果为正面。所以 $P\{T=k,S=1\}=p(1-p)^{n-1}$，而$P(S=1)=\binom{n}{1}p(1-p)^{n-1}$。因此有 $P(T=k|S=1)=P(T=k,S=1)/P(S=1)=1/n$。可见前一看法是正确的。

将上例稍作推广，如果粒子依照参数为 λ 的 Poisson 分布进入计数器，即若记S 为单位时间区间$[0,1]$上的粒子总数则有

$$P(S=k)=\frac{\lambda^k e^{-\lambda}}{k!},\ k=0,1,2\cdots$$

问题是给定在[0，1]时间段上仅有一个粒子到达计数器时该粒子到达时间 T 将服从什么分布？读者很自然会想到是[0，1]上的均匀分布。我们可以把[0，1]区间分成长度为 Δt 的 n 个小区间这时 $\Delta t=\frac{1}{n}$。近似地，好比作了 n 次独立试验粒子以概率 $\lambda\Delta t$ 落入每一小区间，而不在其中的概率为 $1-\lambda\Delta t$。由例 1.11 知

$$P((k-1)\Delta t\leqslant T<k\Delta t|S=1)=\frac{1}{n}=\Delta t.$$

取极限即知给定 $S=1$，T 是[0,1]上的均匀分布。本章习题 10 提供另一解法，不妨一试。

条件期望有些重要的性质我们总结为下面的

命题 1.1 a）若 X 与 Y 独立，则 $E(X|Y=y)=EX$。b)条件期望有所谓平滑性：

$$EX=\int E(X|Y=y)dF_Y(y)=E[E(X|Y)]. \qquad (1.14)$$

c）对随机变量 X，Y 的函数 $\phi(X,Y)$ 恒有

$$E[\phi(X,Y)|Y=y]=E[\phi(X,y)|Y=y]. \qquad (1.15)$$

证明：注意到当 X 与 Y 独立时，给定 $Y=y$ 的 X 的条件分布与无条件分布是一样的。a）由此不难得证。对于 b）仅证明离散随机变量的情形：

$$EX=\sum_k E(X|Y=y_k)P(Y=y_k). \qquad (1.16)$$

由定义 $E(X|Y=y_k)=\sum_j x_jP(X=x_j|Y=y_k)$，再由条件概率的定义(1.16)式右边为 $\sum_k\sum_j x_jP(X=x_j,\ Y=y_k)$先对 k 求和即得$\sum_j x_jP(X=x_j)$由定义正是EX。c)的证明留给读者作为练习。

命题 1.1 b）告诉我们计算 X 的期望可以分两步走，先计算给定 $Y=y$ 时 X 的条件期望。再对这条件期望作加权平均就是总的平均或期望。这是一个很有用的思想。

条件期望是概率统计中极其重要的概念，在均方误差最小准

则下，若要寻找基于 Y 的对 X 的最佳预报函数 $\varphi(\cdot)$，也就是要求出 φ 使 $E[Y-\varphi(Y)]^2$ 达到最小，可以证明所求出的 $\varphi(y)$ 就是 $E(X|Y=y)$。这是一整套回归理论的基础。

1.2.2 矩母函数及生成函数

定义 1.5 随机变量 X 的矩母函数定义为随机变量 e^{tX} 的期望记作 $g(t)$ 即

$$g(t)=E(e^{tX})=\int e^{tX}dF(x). \tag{1.17}$$

矩母函数刻画了随机变量的许多特征，是研究它们特性的重要工具。当矩母函数存在时它唯一地确定了 X 的分布。通过 $g(t)$ 可以求出 X 的各阶矩，即有

$$E[X^n]=g^{(n)}(0), \quad n\geqslant 1 \tag{1.18}$$

其中 $g^{(n)}(t)$ 是 $g(t)$ 的 n 阶导数在 t 的取值。通过在积分号下求导数容易得出(1.18)式的证明，对相互独立的随机变量 X 和 Y，它们和的矩母函数就等于其矩母函数的积：

$$g_{X+Y}(t)=g_X(t)g_Y(t). \tag{1.19}$$

熟记一些常用随机变量分布的矩母函数有助于辨别和刻画不同的随机变量。书末表 1 列出了常用随机变量的分布，矩母函数，均值和方差。

如果随机变量 Y 等于一串独立随机变量 $X_1, X_2, \cdots, X_N$ 的和且下标 N 也是个整数值随机变量。则 Y 是随机变量的随机和。这在实际问题中常常会碰到。比如，N 是在某一时段到达服务台的顾客数，而 X_i 代表第i 个顾客所需要的服务时间则 $Y=\sum_{i=1}^{N}X_i$ 代表总服务时间。这是排队问题中的重要变量。又比如在讨论人口和生物群体生长模型时在某一地区某类生物的总数为 N，X_i 为第 i 个个体的后代数则 Y 就是该群体的总数。下面讨论随机和 Y 的矩母函数与数字特征的计算。

例 1.12 随机和的矩母函数。记 $X_1, X_2, \cdots$ 为一串独立同分布的随机变量，N 为非负整数值随机变量且与 X 序列相独立。Y 为随机和 $\sum_{i=1}^{N} X_i$。$g_Y(t)$ 为其矩母函数。为求 g_Y 先算条件期望

$$
\begin{aligned}
E[\mathrm{e}^{tY} | N=n] &= E[\mathrm{e}^{t\sum_{i=1}^{N} X_i} | N=n] \\
&= E[\mathrm{e}^{t\sum_{i=1}^{n} X_i} | N=n] = E[\mathrm{e}^{t\sum_{i=1}^{n} X_i}] \\
&= [g_X(t)]^n.
\end{aligned}
$$

于是有 $g_Y(t) = E[\mathrm{e}^{tY}] = E\{E[\mathrm{e}^{tY} | N]\} = E[(g_X(t))^N]$ 对 $g_Y(t)$ 关于 t 求导数即有

$$
\begin{aligned}
g'_Y(t) &= E[N(g_X(t))^{N-1} g_X'(t)], \\
g''_Y(t) &= E[N(N-1)(g_X(t))^{N-2}(g'_X(t))^2 \\
&\quad + N(g_X(t))^{N-1} g''_X(t)].
\end{aligned}
$$

将 $t=0$ 代入上面两式得

$$
\begin{aligned}
EY &= E[NE(X)] = ENEX, \\
EY^2 &= EN \operatorname{Var} X + EN^2 E^2 X, \qquad (1.20) \\
\operatorname{Var} Y &= EN \operatorname{Var} X + E^2 X \operatorname{Var} N.
\end{aligned}
$$

如果把 e^{tX} 的数学期望改为 s^X 的期望就可定义生成函数或概率生成函数。

定义 1.6 若 X 为离散随机变量则期望 $E(s^X)$ 为其概率生成函数，记作 $\phi_X(s)$。特别若 $P(X=k)=p_k \quad k=0,1,2,\cdots$ 则 $\phi_X(s) = \sum_{k=0}^{\infty} p_k s^k$。

由定义知 $\Phi_X(s)$ 是以概率 p_k 为系数的幂级数。$\phi_X(s)$ 与 X 的概率分布也是一一对应的，而且有 $\quad p_0 = \phi_X(0)$

$$
p_k = \frac{1}{k!} \frac{\mathrm{d}^k}{\mathrm{d}s^k} \phi_X(s) \Big|_{s=0} \quad k=1,2,\cdots, \qquad (1.21)
$$

与矩母函数的性质类似若 X, Y 为独立随机变量则它们和式的概

率生成函数为

$$\phi_{X+Y}(s)=\phi_X(s)\cdot\phi_Y(s). \tag{1.22}$$

而 X 的期望及有关高阶矩也可由 $\phi(s)$关于 s 的导函数在 $s=1$ 时的值来确定：

$$EX=\phi'_X(s)\big|_{s=1},$$

$$E\{X(X-1)\cdots(X-r+1)\}=\frac{d^r}{ds^r}\phi_X(s)\Big|_{s=1}. \tag{1.23}$$

在随机和的情形，若 X_i 为离散随机变量其概率生成函数为 $\phi_X(s)$ 则对 $Y=\sum_{i=1}^{N}X_i$ 有

$$\begin{aligned}\phi_Y(s)&=E(s^Y)\\&=E\{E[S^Y|N]\}\\&=E\{(\phi_X(s))^N\}=\phi_N(\phi_X(s)).\end{aligned} \tag{1.24}$$

也即 Y 的生成函数是由 N 与 X 的概率生成函数复合而成的。

1.3 收敛性

在随机过程的研究中，还经常会遇到各种意义下的随机变量序列的收敛性问题。下面简要介绍几种收敛性及它们之间的关系。

定义 1.7 设 X_n，$n\geqslant1$ 是一列随机变量，若存在随机变量 X，使对 $\forall\varepsilon>0$，有

$$\lim_{n\to\infty}P(|X_n-X|\geqslant\varepsilon)=0, \tag{1.25}$$

则称随机变量序列 $\{X_n, n\geqslant1\}$ 以概率收敛于 X，记为 $X_n\xrightarrow{p}X$。

如果事件 $\{\omega: \lim\limits_{n\to\infty}(X_n(\omega)-X(\omega))=0\}$的概率为 1，即

$$P(\lim_{n\to\infty}(X_n-X)=0)=1 \tag{1.26}$$

则称随机变量序列$\{X_n, n\geqslant1\}$几乎必然 (almost sure) 收敛于

X，记为 $X_n \to X$，a.s.，也称为随机变量序列以概率 1 收敛于 X。

例 1.13 在 Benoulli 试验中，设每次试验成功的概率为 p，若以 S_n 记 n 次试验中成功的次数，则 $S_n/n \xrightarrow{p} p$。

证 由于 $S_n \sim B(n,p)$，故由 Markov 不等式，对 $\forall \varepsilon > 0$ 有

$$P\left(\left|\frac{S_n}{n}-p\right| \geqslant \varepsilon\right) = P(|S_n - np| \geqslant n\varepsilon) \leqslant \frac{E(S_n - np)^2}{n^2\varepsilon^2}$$

$$= \frac{np(1-p)}{n^2\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \to 0 \quad (n \to \infty)$$

所以 $S_n/n \xrightarrow{p} p$。

定义 1.8 设随机变量 X 和 X_n，$n \geqslant 1$ 都有有限的二阶矩，如果

$$\lim_{n\to\infty} E(X_n - X)^2 = 0 \tag{1.27}$$

则称 X_n 均方收敛于 X，记为 $X_n \xrightarrow{L_2} X$。这儿“均”是平均，“方”是平方。

在例 1.13 中，由于 $E\left(\frac{S_n}{n}-p\right)^2 = \frac{p(1-p)}{n} \to 0 \ (n\to\infty)$，故 $S_n/n \xrightarrow{L_2} p$。

这三种收敛性之间的关系是：均方收敛和几乎必然收敛都蕴含以概率收敛，反过来不必成立；其次均方收敛和几乎必然收敛则互不包含。

习 题 一

1．令 $X(t)$ 为一二阶矩存在的随机过程。试证它是宽平稳的当且仅当 $EX(s)$ 与 $EX(s)X(s+t)$ 都不依赖 s。

2．记 $U_1,\cdots,U_n$ 为在 $(0,1)$ 中均匀分布的独立随机变量。对 $0<t$，$x<1$ 定义

$$I(t,\ x)=\begin{cases}1,\ x\leqslant t,\\0,\ x>t,\end{cases}$$

并记 $X(t)=\frac{1}{n}\sum_{k=1}^{n}I(t,U_k)$，$0\leqslant t\leqslant 1$，这是 $U_1,\ \cdots,U_n$ 的经验分布函数。试求过程 $X(t)$ 的均值和协方差函数。

3．令 Z_1,Z_2 为独立的正态分布随机变量，均值为 0，方差为 σ^2，λ 为实数。定义过程 $X(t)=Z_1\cos\lambda t+Z_2\sin\lambda t$ 试求 $X(t)$ 的均值函数和协方差函数。它是宽平稳的吗？

4．Poisson 过程 $\{X(t),\ t\geqslant 0\}$ 满足 (i) $X(0)=0$ (ii) 对 $t>s$，$X(t)-X(s)$ 服从均值为 $\lambda(t-s)$ 的 Poisson 分布．(iii) 过程是有独立增量的。试求其均值函数和协方差函数。它是宽平稳的吗？

5．$X(t)$ 为习题 4 中的 Poisson 过程。记 $Y(t)=X(t+1)-X(t)$，试求过程 $Y(t)$ 的均值函数和协方差函数。并研究其平稳性。

6．令 Z_1 和 Z_2 是独立同分布的随机变量。$P(Z_1=-1)=P(Z_1=1)=\frac{1}{2}$，记 $X(t)=Z_1\cos\lambda t+Z_2\sin\lambda t$，$t\in\boldsymbol{R}$。试证 $X(t)$ 是宽平稳的，它是严格平稳的吗？

7．试证若 $Z_0,\ Z_1,\ Z_2,\ \cdots$ 为独立同分布随机变量。定义 $X_n=Z_0+Z_1+\cdots Z_n$，则 $\{X_n,\ n\geqslant 0\}$ 是独立增量过程。

8．若 $X_1,X_2,\cdots$ 为独立随机变量，还要添加什么条件才能确保它是严格平稳的随机过程？

9．令 X 和 Y 是从单位圆上的均匀分布中随机选取一点所得的横坐标和纵坐标。试计算条件概率

$$P\left(X^2+Y^2\geqslant\frac{3}{4}\middle| X>Y\right)$$

10．粒子依参数为 λ 的 Poisson 分布进入计数器，两粒子到达的时间间隔 $T_1,T_2,\cdots$ 是独立的参数为 λ 的指数分布随机变量。记 S 是 $[0,1]$ 时段中的粒子总数。时间区间 $I\subset[0,1]$ 其长度记为

$|I|$。试证明 $P(T_1\in I, S=1)=P(T_1\in I,\ T_1+T_2>1)$ 并由此计算 $P(T_1\in I|S=1)=|I|$。

11．X，Y 为两独立随机变量且分布相同。证明 $E(X|X+Y=z)=E(Y|X+Y=z)$。并试求基于 $X+Y=z$ 的 X 的最佳预报。并求出预报误差 $E(X-\varphi(X+Y))^2$。

12．气体分子的速度 V 有三个垂直分量。V_x, V_y, V_z。它们的联合分布密度依 Maxwell-Boltzman 定律为

$$f_{V_x,\ V_y,\ V_z}(v_x,v_y,v_z)=\frac{1}{(2\pi kT)^{3/2}}\exp-\left(\frac{-v_x^2+v_y^2+v_z^2}{2kT}\right),$$

其中 k 是 Boltzman 常数。T 为绝对温度。给定分子的总动能为 e。试求 x 方向的动量的绝对值的期望值。

13．若 $X_1, X_2, \cdots, X_n$ 独立同分布。它们服从参数为 λ 的指数随机变量。试证 $\sum_{i=1}^{n}X_i$ 是参数为 $(n,\ \lambda)$ 的 Γ 分布。其密度为

$$f(t)=\lambda e^{-\lambda t}(\lambda t)^{n-1}/(n-1)!,\quad t\geqslant 0。$$

14．设 X_1 和 X_2 为相互独立的均值为 λ_1 和 λ_2 的 Poisson随机变量试求 X_1+X_2 的分布，并计算给定 $X_1+X_2=n$ 时 X_1 的条件分布。

15．若 $X_1, X_2, \cdots$ 独立且有相同的以 λ 为参数的指数分布。N 服从几何分布，即

$$P(N=n)=\beta(1-\beta)^{n-1},\quad n=1,2,\cdots,\quad 0<\beta<1。$$

试求随机和 $Y=\sum_{i=1}^{N}X_i$ 的分布。

16．若 $X_1, X_2, \cdots$ 独立同分布 $P(X_i=\pm 1)=\frac{1}{2}$ N 与 X_i 独立且服从参数为 β 的几何分布，$0<\beta<1$。试求随机和 $Y=\sum_{i=1}^{N}X_i$ 的均值，方差和三，四阶矩。

17．随机变量 N 服从参数为 λ 的 Poisson 分布。给定 $N=n$，随机变量 M 服从以 n 和 p 为参数的二项分布。试求 M 的无条件概率分布。

第二章 Poisson过程

2.1 Poisson 过程

许多偶然现象可以用 Poisson 分布来描述。大量自然界的物理过程可以用 Poisson 过程来刻画。它是随机建模的重要基石，也是学习随机过程理论的重要直观背景。最著名的例子包括盖格计数器上的粒子流，二次大战时伦敦空袭的弹着点，电话总机所接收到传呼的次数，交通流中的事故数，地震记录，细胞中染色体的交换等。这类变化过程可粗略地假定为有相同的变化类型。我们所关心的是随机事件的数目，而每一变化可用时间或空间上的一个点来表示。这类过程有如下两个性质：一是在时间或空间上的均匀性，二是未来的变化与过去的变化没有关系。我们将基于这些性质推导出 Poisson 过程的模型。

定义 2.1 一个整数值随机过程 $\{N(t),\ t\geqslant 0\}$ 满足下述三个条件就称作强度为 $\lambda>0$ 的 Poisson 过程。(i) $N(0)=0$; (ii) $N(t)$ 是独立增量过程；(iii) 对任何 $t>0$，$s\geqslant 0$ 增量 $N(s+t)-N(s)$ 服从参数为 λt 的 Poisson 分布，即

$$P\{N(s+t)-N(s)=k\}=\frac{(\lambda t)^k e^{-\lambda t}}{k!},\quad k=0,1,\cdots. \tag{2.1}$$

定义的条件 (i) 告诉我们随机事件从时刻 0 开始计数。条件 (iii) 是过程称为 Poisson 过程的直接理由。由 Poisson 分布的性质我们马上可以求出 $EN(t)=\mathrm{Var}[N(t)]=\lambda t$. 增量 $N(s+t)-N(s)$ 代表时间区间 $(s,s+t]$ 中发生的随机事件数，由条件 (iii) 这一增量的分布与 s 无关因而增量具有平稳性。因此条件 (ii) 和 (iii)

充分刻画了模型所要求的两个性质：前后的独立性与时间上的均匀性。强度 λ 有时也称为速率。它描绘随机事件发生的频繁程度。

例 2.1　顾客依 Poisson 过程到达某商店，速率为 $\lambda=4$ 人/小时。已知商店上午 9:00 开门。试求到 9:30 时仅到一位顾客，而到 11:30 时总计已到达 5 位顾客的概率。

解　令 t 的计时单位为小时，并以 9:00 为起始时刻，所求事件可表示为 $\left\{N\left(\frac{1}{2}\right)=1, N\left(\frac{5}{2}\right)=5\right\}$。其概率为

$$
\begin{aligned}
&P\left\{N\left(\frac{1}{2}\right)=1,\ N\left(\frac{5}{2}\right)=5\right\} \\
&=P\left\{N\left(\frac{1}{2}\right)=1,\ N\left(\frac{5}{2}\right)-N\left(\frac{1}{2}\right)=4\right\} \\
&=\left(\frac{e^{-4\cdot\frac{1}{2}}\cdot 4\cdot\frac{1}{2}}{1!}\right)\left(\frac{e^{-4\cdot 2}(4\cdot 2)^4}{4!}\right) \\
&=0.0155.
\end{aligned}
$$

稀有事件的概率常服从 Poisson 分布。这是由于当试验次数很多而每次试验成功的概率很小时，二项分布可以逼近 Poisson 分布。若记 N 为试验次数，p 为成功概率而当 N 很大时 Np 趋于一个常量 λ 则 N 次试验中的总成功次数近似地服从参数为 λ 的 Poisson分布。这一想法很自然地可以推广到随机过程的情况。记 $[0,\infty)$ 为观察过程的时间轴，0 代表起始时刻，$N(b)-N(a)$ 代表时间区间 $(a,b]$ 上发生的事件数。我们特作如下假定：

(1) 在不相交区间中事件发生的数目相互独立，也即对任何整数 $n=2,3,\cdots$ 如时刻 $t_0=0<t_1<t_2<\cdots<t_n$，增量 $N(t_1)-N(t_0)$，$N(t_2)-N(t_1)$，$\cdots$，$N(t_n)-N(t_{n-1})$ 相互独立。

(2) 对任何时刻 t 和正数 h，随机变量（增量）$N(t+h)-N(t)$ 的概率分布只依赖于区间长度 h 而不依赖时刻 t。

(3) 存在正常数 λ，当 $h\downarrow 0$ 时，使在长度为 h 的小区间中事

件至少发生一次的概率

$$P\{N(t+h)-N(t)\geq 1\}=\lambda h+o(h)$$

(4) 在小区间$(t,t+h]$发生两个或两个以上事件的概率为$o(h)$(可以忽略不计),即当$h\downarrow 0$,

$$P\{N(t+h)-N(t)\geq 2\}=o(h)$$

从逼近的观点看(1)说明试验是独立的;(2)说明在每个长度相同的小区间上事件发生有相同的概率p;(3)告诉我们成功(事件发生)概率$p\doteq\lambda h$,而且p很小;(4)是说明事件不发生的概率为$1-\lambda h\doteq 1-p$。这正好是独立Bernoulli试验的模型。若观察区间为$[0,t]$则$N\doteq\frac{t}{h}$。所以当$h\downarrow 0$时,$N\to\infty$。而$Np\doteq\lambda t$。二项分布的极限为参数为λt的Poisson分布。从直观意义看如前所述(1)为前后的独立性,(2)为时间上的均匀性或齐次性,(3)表明事件是稀有的,而(4)则称为相继性(orderliness)。意思是事件是一件一件地发生的,在同一瞬间同时发生多个事件的可能性很小很小。基于这些假定我们可以证明。

命题 2.1 满足假定(1)—(4)的随机过程$N(t)$为Poisson过程.

证明 由假定(1)和(2),$N(s+t)-N(s)$的分布和$N(t)$的分布相同。为验证过程$N(t)$是Poisson过程,只需求出$N(t)$的概率分布。记$P_m(t)=P\{N(t)=m\}$,并记$p(h)=P\{N(h)\geq 1\}=P_1(h)+P_2(h)+\cdots=1-P_0(h)$。$p(h)$是在$(0,h]$上发生一个及一个以上事件的概率.由独立性

$$P_0(t+h)=P_0(t)P_0(h)=P_0(t)(1-p(h))。$$

因而,

$$\frac{P_0(t+h)-P_0(t)}{h}=-P_0(t)\frac{p(h)}{h}.$$

令$h\downarrow 0$,由假设(3)得到微分方程$P_0'(t)=-\lambda P_0(t)$。容易看出这一方程的解是$P_0(t)=Ce^{-\lambda t}$。由初始条件$P_0(0)=1$确定$C=1$。

故有 $P_0(t)=e^{-\lambda t}$。下面再依次定出 $P_i(t)$，$i\geqslant 1$。由独立增量性，

$$P_1(t+h)=P_1(t)P_0(h)+P_0(t)P_1(h). \qquad (2.2)$$

由假定(4)可推出

$$\begin{aligned} p(h) &= P\{N(t+h)-N(t)\geqslant 1\} \\ &= P_1(h)+P\{N(t+h)-N(t)\geqslant 2\} \\ &= P_1(h)+o(h). \end{aligned}$$

将 $P_0(h)=1-p(h)$ 及 $P_1(h)=p(h)+o(h)$ 代入(2.2)后再稍加整理有

$$P_1(t+h)-P_1(t)=-p(h)P_1(t)+p(h)P_0(t)+o(h)P_0(t).$$

两边除以 h 并令 $h\downarrow 0$ 可得到微分方程

$$P_1'(t)=-\lambda P_1(t)+\lambda P_0(t).$$

这里所用的分析方法是仅让 t 变化很小利用已知假设建立起有关概率的微分方程，常称为无穷小分析。类似的对 $m>1$ 有

$$\begin{aligned} P_m(t+h) &= P_m(t)P_0(h)+P_{m-1}(t)P_1(h) \\ &\quad +\sum_{i=2}^{m}P_{m-i}(t)P_i(h). \end{aligned} \qquad (2.3)$$

注意到，由假设(4)

$$\sum_{i=2}^{m}P_{m-i}(t)P_i(h)\leqslant\sum_{i=2}^{\infty}P_i(h)=o(h),$$

将(2.3)稍加整理并令 $h\downarrow 0$ 即有

$$P_m'(t)=-\lambda P_m(t)+\lambda P_{m-1}(t),\quad m=1,2,\cdots. \qquad (2.4)$$

在初始条件 $P_m(0)=0$，$m=1,2,\cdots$ 下容易解出(习题 5)

$$P_m(t)=\frac{\lambda^m t^m}{m!}e^{-\lambda t}.$$

由定义 2.1 命题得证。

不直接解方程(2.4)，也可利用 1.2.2 节所介绍的矩母函数方法来验证 $N(t)$ 的矩母函数与 Poisson 分布的矩母函数相同；记

$$g(u,t)=g_{N(t)}(u)=Ee^{uN(t)}=\sum_{k=0}^{\infty}e^{uk}P_k(t)$$

为过程 $N(t)$ 的矩母函数。由方程(2.4)知

$$\frac{\partial g(u,t)}{\partial t}=\sum_{k=0}^{\infty}e^{uk}P_k'(t)$$

$$=-\lambda\sum_{k=0}^{\infty}e^{uk}P_k(t)+\lambda\sum_{k=1}^{\infty}e^{uk}P_{k-1}(t)$$

$$=-\lambda g(u,t)+\lambda e^u g(u,t).$$

所以立即有

$$\begin{cases}\dfrac{\partial}{\partial t}\ln g(u,t)=\lambda(e^u-1),\\ g(u,o)=P_0(0)=1.\end{cases}$$

解出 $g(u,t)=e^{\lambda t(e^u-1)}$，正是 Poisson 变量当参数为 λt 时的矩母函数。

命题 2.1 的这些假定在实际问题中常常可以近似地得到满足，从而建立起 Poisson 过程的随机模型。

2.2 与Poisson过程相联系的若干分布

我们考虑直线上的 Poisson 过程 $N(t)$。过程的一次实现或样本路径一般为跳跃度为 1 的阶梯函数如图 2.1 所示。这个图对理解和掌握 Poisson 过程十分有用。

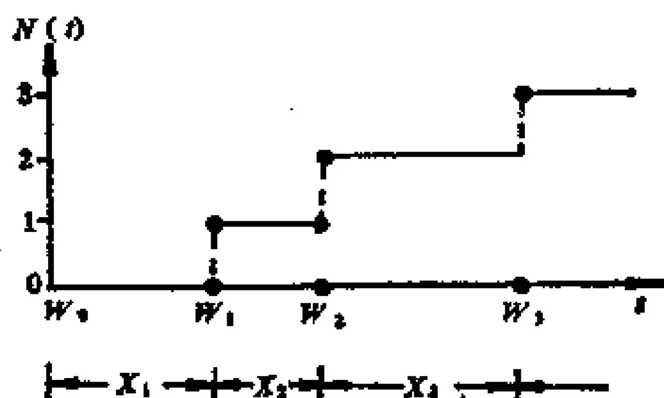


图 2.1 Poisson 过程的样本路径

图中第 $n-1$ 次与第 n 次事件间的间隔时间记作 X_n。而 $W_n=\sum_{i=1}^{n}X_i$ 为第 n 次事件的到达或等待时间。我们来求 X_n 与 W_n 的

分布。

命题 2.2 $X_n, n=1,2,\cdots$ 是均值为$\frac{1}{\lambda}$的独立同分布的指数随机变量。W_n 服从参数为 n 和 λ 的 Γ 分布。

证明 事件 $\{X_1>t\}$表示第 1 次事件发生在时刻 t 之后。其发生当且仅当在时间区间 $(0,t]$ 中 Poisson 过程不曾有事件发生过。所以，

$$P\{X_1>t\}=P\{N(t)=0\}=\mathrm{e}^{-\lambda t}.$$

而

$$\begin{aligned} P\{X_2>t|X_1=s\}&=P\{(s,s+t]\text{ 中事件不发生}|X_1=s\}\\ &=P\{(s,s+t]\text{中事件不发生}\} \quad \text{（独立增量性）}\\ &=P\{(0,t]\text{ 中事件不发生}\} \quad \text{（平稳增量性）}\\ &=\mathrm{e}^{-\lambda t}. \end{aligned}$$

这就可以得出结论 X_2 也服从以 λ 为参数的指数分布而且 X_1 与 X_2 是独立的。类似地可对其他 X_i 证明命题的结论。利用第一章习题 13 可以知道 W_n 服从 Γ 分布，参数为 (n, λ)。但我们宁愿在这里再给一个直接的证明。因事件 $\{N(t)\geqslant n\}$ 是与$\{W_n\leqslant t\}$等价的。它们都表明第 n 次事件发生在时刻 t 之前或者换言之到时刻 t 已经至少发生了 n 件事。于是

$$P\{W_n\leqslant t\}=P\{N(t)\geqslant n\}=\sum_{j=n}^{\infty}\mathrm{e}^{-\lambda t}\ \frac{(\lambda t)^j}{j!}. \tag{2.5}$$

对 W_n 的分布函数(2.5)式关于 t 求导即可求出 W_n 的密度函数

$$\begin{aligned} f_{W_n}(t)&=-\sum_{j=n}^{\infty}\lambda\mathrm{e}^{-\lambda t}\frac{(\lambda t)^j}{j!}+\sum_{j=n}^{\infty}\lambda\mathrm{e}^{-\lambda t}\frac{(\lambda t)^{j-1}}{(j-1)!}\\ &=\lambda\mathrm{e}^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!}. \end{aligned}$$

命题遂得证。

命题2.2从另一角度刻画了Poisson过程而且可以应用于对过程进行计算机模拟。从均匀分布 $U[0,1]$ 中随机抽样的方法是熟知的。将其作变换后就可以模拟参数为 λ 的指数随机变量。而独立同分布的指数随机变量的和$\sum_{i=1}^{n}X_i$ 依命题 2.2 就是 W_n，也就是

Poisson 过程第 n 次事件到达的时刻。这样就可以模拟参数为 λ 的 Poisson 过程 $\{N(t),t\geqslant 0\}$。λ 越大表明事件平均间隔时间 $\frac{1}{\lambda}$ 越短，事件的发生就越频繁，强度也就越大。

给定了到时刻 t 总计发生了 n 件事，在过去某时刻 $u<t$ 发生了 k 件事的条件概率可由习题 1 知道是二项分布。但人们常常对给定 $N(t)=n$ 后事件发生在 $W_1,\cdots,W_n$ 的联合分布更感兴趣。关于这一条件联合分布，我们有

定理 2.1 若 $N(t)$，$t\geqslant 0$ 为 Poisson 过程。给定 $N(t)=n$ 等待时间 $W_1,\cdots,W_n$ 的联合密度为

$$f_{W_1,\cdots,W_n|N(t)=n}(w_1,\cdots,w_n|n)=n!t^{-n},\quad 0<w_1<\cdots<w_n\leqslant t. \tag{2.6}$$

证明 给定 $N(t)=n$，不妨设 n 个等待时间 $W_i=w_i$，$i=1,\cdots,n$。对充分小的增量 Δw_i，事件 $N(t)=n$ 和 $w_i\leqslant W_i<w_i+\Delta w_i$，$i=1,\cdots,n$ 是与在 $[w_i,w_i+\Delta w_i)$，$i=1,\cdots,n$ 中恰恰发生了一件事而在 $[0,w_1)$，$[w_1+\Delta w_1,w_2)$，$\cdots$，$[w_{n-1}+\Delta w_{n-1},w_n)$，$[w_n+\Delta w_n,t]$ 中没有发生事件是相对应的。于是

$$\begin{aligned}
&f_{W_1,\cdots,W_n|N(t)=n}(w_1,\cdots,w_n|n)\Delta w_1,\cdots,\Delta w_n\\
&\quad=P\{w_i\leqslant W_i<w_i+\Delta w_i,i=1,2,\cdots,n|N(t)=n\}\\
&\qquad+o(\Delta w_1\cdots\Delta w_n)\\
&\quad=P\{w_i\leqslant W_i<w_i+\Delta w_i,i=1,2,\cdots,n;N(t)=n\}\\
&\qquad/P\{N(t)=n\}+o(\Delta w_1\cdots\Delta w_n).
\end{aligned} \tag{2.7}$$

上式第一项的分子由独立增量性及 Poisson 过程的定义知为

$$\begin{aligned}
&P\{N(w_i+\Delta w_i)-N(w_i)=1,i=1,2,\cdots,n;\\
&\quad N(w_1)=0,N(w_2)-N(w_1+\Delta w_1)=0,\cdots,\\
&\quad N(w_n)-N(w_{n-1}+\Delta w_{n-1})=0,N(t)-N(w_n+\Delta w_n)=0\}\\
&=[\lambda\Delta w_1\cdots\lambda\Delta w_n+o(\Delta w_1\cdots\Delta w_n)]\cdot\\
&\quad[e^{-\lambda w_1}e^{-\lambda(w_2-w_1-\Delta w_1)}\cdots e^{-\lambda(w_n-w_{n-1}-\Delta w_{n-1})}e^{-\lambda(t-w_n-\Delta w_n)}]\\
&=\lambda^n\Delta w_1\cdots\Delta w_n\cdot e^{-\lambda t}\cdot e^{\lambda\sum_{i=1}^{n}\Delta w_i}+o(\Delta w_1\cdots\Delta w_n)
\end{aligned}$$

分母为 $e^{-\lambda t}(\lambda t)^n/n!$。代入后在(2.7)式两边除以 $\Delta w_1\cdots\Delta w_n$ 并令 $\Delta w_i\to 0$ $i=1,2,\cdots,n$ 就得到(2.6)式。

熟悉数理统计的读者马上可以看出联合密度(2.6)似曾相识。若从 $[0,t]$ 区间上的均匀分布抽取 n 个独立同分布的随机样本 $U_1,\cdots,U_n$ 并按其大小排列成次序统计量记为 $U_{(1)}<U_{(2)}<\cdots<U_{(n)}$。则 $U_{(i)}$, $i=1,\cdots,n$ 的联合密度正好是(2.6)。有关证明可参看陈希孺编的《数理统计学教程》，上海科技出版社，1988，33页。

例 2.2 顾客依速率为 λ 的 Poisson 过程到达车站。若火车在时刻 t 离站，问在 $(0,t]$ 区间里顾客的平均总等待时间是多少？

解 作为依 Poisson 过程到达的第一位顾客他的到达时间为 W_1。等到时刻 t 发车需等待 $t-W_1$。而第 i 位旅客的等待时间为 $t-W_i$。在 $(0,t]$ 区段总共来了 $N(t)$ 位客人，所以总等待时间为 $\sum_{i=1}^{N(t)}(t-W_i)$。而所要求的平均总等待时间就是 $E\left[\sum_{i=1}^{N(t)}(t-W_i)\right]$。为求出它可以先求条件期望

$$E\left[\sum_{i=1}^{N(t)}(t-W_i)\,|\,N(t)=n\right]$$

$$=E\left[\sum_{i=1}^{n}(t-W_i)\,|\,N(t)=n\right]$$

$$=nt-E\left[\sum_{i=1}^{n}W_i\,|\,N(t)=n\right].$$

注意到给定 $N(t)=n$，W_i, $i=1,\cdots,n$ 的联合密度是与 $(0,t]$ 上均匀分布中随机样本 U_i, $i=1,\cdots,n$ 的次序统计量 $U_{(i)}$, $i=1,\cdots,n$ 的联合密度是一样的。于是，

$$E\left[\sum_{i=1}^{n}W_i\,|\,N(t)=n\right]$$

$$=E\left[\sum_{i=1}^{n}U_{(i)}\right]=E\left[\sum_{i=1}^{n}U_i\right]=\frac{nt}{2}.$$

因此，

$$E\left[\sum_{i=1}^{N(t)}(t-W_i)\,|\,N(t)=n\right]=nt-\frac{nt}{2}=\frac{nt}{2}.$$

最后得到

$$E\left[\sum_{i=1}^{N(t)}(t-W_i)\right]=\frac{t}{2}E[N(t)]=\frac{\lambda t^2}{2}.$$

可以看到顾客平均总等待时间是和 t^2 成正比的，比例因子的大小决定于 Poisson 过程的强度 λ。

2.3 Poisson过程的推广

前两节所讨论的只是 Poisson 过程大家庭中最简单，最基本的一种情形，是大大简化了的随机模型。人们可以很自然地将其作各种推广。

2.3.1 非齐次 Poisson 过程

Poisson 过程 $N(t)$ 中的强度 λ 是常数。它是事件发生在某一小区间上的概率与区间长度的比例因子。更确切地说

$$P\{N(t+h)-N(t)=1\}=\lambda h\cdot e^{-\lambda h}=\lambda h+o(h)$$

如果允许这一比例因子依赖于时刻 t，所得的过程就是非齐次 Poisson 过程。其参数为 t 的函数 $\lambda(t)$。这也提供了过程增量不平稳的例子。此时，

$$P\{N(t+h)-N(t)=k\}$$
$$=\frac{\left(\int_t^{t+h}\lambda(u)\,du\right)^k\exp\left(-\int_t^{t+h}\lambda(u)\,du\right)}{k!},\quad k=0,1,\cdots.\tag{2.8}$$

人们只须将 2.1 节中的假定(3)改为(3)′：

$$P\{N(t+h)-N(t)\geqslant 1\}=\lambda(t)h+o(h)$$

就可以完全类似地证明。

命题 2.3 在 2.1 节的假定(1)，(2)，(4)和本节假定(3)下过

程 $N(t)$ 是参数为 $\lambda(t)$ 的非齐次Poisson 过程，过程增量$N(t+h)-N(t)$ 的分布为(2.8)式。

例 2.3 记录值。随机变量 $X_1, X_2, \cdots$ 是一串独立同分布连续随机变量。分布为 $F(t)$，密度函数为 $f(t)$。当 X_i 代表某一元件的寿命时 $\lambda(t)=f(t)/(1-F(t))$常定义为是失效率。它的直观意义是 $\lambda(t)$ 近似地等于 $P(X_1=t|X_1>t)$，即元件在时刻 t 时仍在工作而在下一瞬间 t 失效的条件概率。记 $X_0=0$，当 $X_n>\max(X_1,\cdots,X_{n-1})$时，也就是当 X_n 的值超过以往任何记录时称在时刻 n 创了记录。X_n 则称为是所创的记录值。记录值比 t 小的新记录的次数记为 $N(t)$。也即若记录值为 x 则事件在 x 处发生则 $\{N(t), t\geq 0\}$ 为一非齐次 Poisson 过程。其强度函数正好是失效率 $\lambda(t)$。比如，

$$
\begin{aligned}
P(N(t)=0) &= P(X_1>t)=1-F(t)=\mathrm{e}^{\ln(1-F(t))} \\
&= \exp\left[-\int_0^t \frac{\mathrm{d}F(u)}{1-F(u)}\right]=\exp\left[-\int_0^t \frac{f(u)\mathrm{d}u}{1-F(u)}\right] \\
&= \frac{\left(\int_0^t \lambda(u)\mathrm{d}u\right)^k}{k!}\exp\left[-\int_0^t \lambda(u)\mathrm{d}u\right]\Bigg|_{k=0}.
\end{aligned}
$$

正好是(2.8)式中 $k=0$ 的情形。有兴趣的读者可自行证明本例。记录值的模型也解释了为什么常用 Poisson 过程来描述事故与灾害的发生过程，因为灾害就是负面的破记录。

2.3.2 复合 Poisson 过程

人们在考虑设备故障所需的维修费，自然灾害所造成的损失，股票市场的价格变动时都会碰到这样一类模型：事件的发生依从一 Poisson 过程，而每一次事件都还附带一个随机变量（如费用，损失等）。这时，人们感兴趣的不仅仅是事件发生的次数，人们还要了解总费用或总损失。这也就是累计值过程 $X(t)=\sum_{i=1}^{N(t)} Y_i$，其中 Y_i 为独立同分布的随机变量。它们有分布函数 $G(y)$，均值

Poisson过程.

$EY=\mu$，方差 $\mathrm{Var}Y=\tau^2$. $N(t)$ 是参数为 λ 的 Poisson 过程. $X(t)$ 就是复合 Poisson 过程。由例 1.12 易知 $X(t)$ 是随机和并且 $E[X(t)]=\lambda\mu t$，$\mathrm{Var}[X(t)]=\lambda(\tau^2+\mu^2)t$. 若取 Y_i 恒等于 1，复合 Poisson 过程 $X(t)$ 就是通常的 Poisson 过程.

例 2.4 假定在股票交易市场。股票交易次数为以 λ 为速率的 Poisson 过程。记第 k 次与第 $k-1$ 次易手前后股票价格的变化为 Y_k. 不妨假定 $Y_1, Y_2, \cdots$ 是独立同分布的随机变量且与 $N(t)$ 独立。而 $X(t)=\sum_{k=1}^{N(t)} Y_k$ 则代表到时刻 t 时股票的总价格变化。这是投资者计算赢亏决定投资意向的重要指示。设 Y_1 的分布为 $G(y)$. 则 $P\{Y_1+Y_2\leqslant y\}=\int_{-\infty}^{\infty} G(y-z)\mathrm{d}G(y)$ 记为 $G*G(y)$ 或 $G^{(2)}(y)$ 是 G 与 G 自身的卷积。类似的 $G^{(n)}(y)=G*\cdots*G(y)$ 记 G 自身的 n 重卷积。基于上述记号，不难求出 $X(t)$ 的分布：

$$
\begin{aligned}
P\{X(t)\leqslant x\} &= P\left\{\sum_{k=1}^{N(t)} Y_k \leqslant x\right\} \\
&= \sum_{n=0}^{\infty} P\left\{\sum_{k=1}^{N(t)} Y_k \leqslant x \mid N(t)=n\right\} P\{N(t)=n\} \\
&= \sum_{n=0}^{\infty} P\left\{\sum_{k=1}^{n} Y_k \leqslant x\right\} \frac{(\lambda t)^n \mathrm{e}^{-\lambda t}}{n!} \\
&= \sum_{n=0}^{\infty} \frac{(\lambda t)^n \mathrm{e}^{-\lambda t}}{n!} G^{(n)}(x).
\end{aligned}
$$

2.3.3 标值(Marked) Poisson 过程

在复合 Poisson 过程中我们关心的是累计值。若将发生事件的时刻和相伴的随机变量值一并考虑，引进随机对或随机向量 $(W_1, Y_1), \cdots, (W_k, Y_k), \cdots$ 则得到所谓标值 Poisson 过程。其中 W_k 仍为第 k 个事件的等待时间。这是一个多维的过程，限于篇幅不能作详细的讨论。但随着对独立同分布随机变量的不同定义，人们可以讨论各种有用的模型。比如令 $P\{Y_k=1\}=p$，$P\{Y_k=0\}=q=1-p$，$0<p<1$. 就可以得到本章习题 9 所讨论

的模型。再令 $N_1(t)=\sum_{k=1}^{N(t)} Y_k$ 则 $N_1(t)$ 是强度为 λp 的Poisson过程。它是以一定的概率机制从 $N(t)$ 抽取的。这一机制称为稀疏（thinning）。

2.3.4 空间 Poisson 过程

直线上的 Poisson 过程可以自然地推广到空间去。记 S 为 n 维空间的集合，$\mathscr{A}$ 为 S 的子集所组成的集合类，其中元素为 $A,B,\cdots$。$N(A)$ 为一随机过程，它的取值是非负整数值 $\{0,1,2,\cdots\}$ 它的指标集是 $\mathscr{A}$ 中集合 A。我们最基本的要求是 $N(\cdot)$ 有某种可加性。即如果 A 与 B 不相交则 $N(A\cup B)=N(A)+N(B)$。回想在一维的情形 S 就是正半实轴。$A=(s,t]$，$0\leqslant s<t$。我们记 A 的大小为 $|A|$（长度、面积、体积等）。

定义 2.2 $\{N(A),A\in\mathscr{A}\}$ 若满足下面两个条件就称为强度是 $\lambda>0$ 的空间齐次 Poisson 过程

(i) 对每个 $A\in\mathscr{A}$，$N(A)$ 服从参数为 $\lambda|A|$ 的 Poisson 分布。

(ii) 对任何有限的由 S 的不交子集组成的集合类 $\{A_1,\cdots,A_n\}$，随机变量 $N(A_1),\cdots,N(A_n)$ 是相互独立的。

例如在天文学中常常可以把宇宙空间星星的分布看成三维的Poisson过程，而二次大战空袭时伦敦的弹着点数是二维Poisson过程的著名例子。

2.3.5 更新过程

由命题 2.2 知事件的间隔时间 X_k 相互独立而且有相同的均值为 $\frac{1}{\lambda}$ 的指数分布。这是 Poisson 过程的重要特征。如果把时间间隔 X_k 服从的指数分布改为一般的分布函数 $F(x)$，那么所得的将会是什么过程呢？这就是所谓更新过程。

定义 2.3 如果 X_i，$i=1,2,\cdots$ 为一串非负的随机变量。它们独立同分布，分布函数为 $F(x)$。记 $W_0=0$，$W_n=\sum_{i=1}^{n}X_i$。W_n 表示第 n 次事件发生的时刻，则称

$$N(t)=\max\{n:W_n\leqslant t\}$$

为更新过程。

定义中 $N(t)$ 代表了到时刻 t 时事件的总数。$W_i,i=1,2,\cdots$ 也常常称为是更新点。在这些更新点上过程又重新开始。在更新过程中事件发生的平均次数称为是更新函数。记作 $m(t)$。即 $m(t)=E[N(t)]$。更新理论的主体是研究更新函数的性质。我们仅给出最基本的。

命题 2.3 更新过程 $N(t)$ 的分布

$$P\{N(t)=n\}=F^{(n)}(t)-F^{(n+1)}(t),$$

而更新函数 $m(t)=\sum_{n=1}^{\infty}F^{(n)}(t)$。其中 $F^{(n)}(t)$ 为 $F(t)$ 的 n 重卷积，$F(t)$ 即为 X_i 的分布函数。

证明 首先不难看出到时刻 t 的更新总数大于或等于 n 是与第 n 次更新发生在时刻 t 前是等价的。也即有

$$\{N(t)\geqslant n\}\Longleftrightarrow\{W_n\leqslant t\}.$$

于是

$$\begin{aligned}P\{N(t)=n\}&=P\{N(t)\geqslant n\}-P\{N(t)\geqslant n+1\}\\&=P\{W_n\leqslant t\}-P\{W_{n+1}\leqslant t\}=F^{(n)}(t)-F^{(n+1)}(t).\end{aligned}\tag{2.9}$$

为求出更新函数，引入示性函数

$$I_n=\begin{cases}1, & \text{若第 } n \text{ 次更新发生在 } [0,t] \text{ 中},\\0, & \text{如若不然},\end{cases}$$

显然，$N(t)=\sum_{n=1}^{\infty}I_n$ 于是

$$\begin{aligned}E[N(t)]&=E\left[\sum_{n=1}^{\infty}I_n\right]=\sum_{n=1}^{\infty}E[I_n]\\&=\sum_{n=1}^{\infty}P\{I_n=1\}=\sum_{n=1}^{\infty}P\{W_n\leqslant t\}=\sum_{n=1}^{\infty}F^{(n)}(t).\end{aligned}\tag{2.10}$$

例 2.5 对 Poisson 过程 $N(t)$，$X_1, X_2, \cdots$ 为独立同指数分布。由 2.2 节中的(2.5)式知

$$F^{(n)}(t)=\sum_{j=n}^{\infty} e^{-\lambda t}\frac{(\lambda t)^j}{j!},$$

这也可以由对 Γ 分布的密度 $f_{W_n}(t)$ 求积分来直接证明，于是利用命题 2.3 立即可得

$$\begin{aligned}P\{N(t)=n\}&=F^{(n)}(t)-F^{(n+1)}(t)\\&=\sum_{j=n}^{\infty}e^{-\lambda t}\frac{(\lambda t)^j}{j!}-\sum_{j=n+1}^{\infty}e^{-\lambda t}\frac{(\lambda t)^j}{j!}=e^{-\lambda t}\frac{(\lambda t)^n}{n!}\end{aligned}$$

而

$$\begin{aligned}m(t)&=\sum_{n=1}^{\infty}F^{(n)}(t)=\sum_{n=1}^{\infty}\sum_{j=n}^{\infty}e^{-\lambda t}\frac{(\lambda t)^j}{j!}\\&=\sum_{j=1}^{\infty}\sum_{n=1}^{j}\left(\frac{e^{-\lambda t}(\lambda t)^j}{j!}\right)=\sum_{j=1}^{\infty}\frac{e^{-\lambda t}(\lambda t)^j}{(j-1)!}\\&=\lambda t\sum_{k=0}^{\infty}\frac{e^{-\lambda t}(\lambda t)^k}{k!}=\lambda t.\end{aligned}$$

这些结果都是我们所熟知的了。Poisson 过程确实为更新过程的特殊情形。

更新理论有着丰富的内容和广泛的应用。运筹学中的排队论，存贮理论，可靠性数学。管理数学都要用到大量的更新理论的知识。在进一步学习这些内容时经常回想到 Poisson 过程是它的特例对于加深理解颇有益处。比如对 Poisson 过程 $EN(t)/t=\lambda t/t=\lambda=1\Big/\left(\frac{1}{\lambda}\right)$恒为常数。而令人惊奇的是对一般的更新过程也有$\lim\limits_{t\to\infty}\frac{EN(t)}{t}=\frac{1}{\mu}$，其中$\mu$是事件间隔时间 X_i 的期望 EX_i。这就是著名的基本更新定理。它告诉我们单位时间的平均事件次数当 t 趋于无穷时趋于一个固定的极限$\frac{1}{\mu}$。

我们已从上述五个方面推广了过程理论中的一块重要基石 Poisson 过程。在学习随机过程理论和处理各种随机模型时应注意它们的联系和区别。要弄清所研究的实际问题的适用条件是什

么。这样才能更准确地刻画实际的物理过程。在下一章 Markov 过程的学习之后我们还可以看到 Poisson 过程是一类非常特殊的连续时间 Markov 过程——纯生过程的特例。Poisson 过程随时间的推移，每次事件发生时 $N(t)$ 增加 1，随后，经过以相同参数 λ 的指数分布等待时间后事件再次发生过程再增加 1。而对一般的纯生过程，过程的状态变化的等待时间可以是不同参数的指数分布。至于更一般的生灭过程，跳跃度甚至可以是负数。尽管问题复杂化了，但本章命题 2.1 中所用到的分析方法却仍然广泛适用。

习 题 二

1．$N(t)$ 为一Poisson 过程，对 $s<t$ 试求条件概率 $P\{N(s)=k|N(t)=n\}$。

2．$\{N(t),t\geqslant 0\}$ 为一强度是 λ 的 Poisson 过程。对 $s>0$，试计算 $E[N(t)\cdot N(t+s)]$。

3．电报依平均速率为每小时 3 个的 Poisson 过程到达电报局。试问

(i) 从早上八时到中午没有收到电报的概率？

(ii) 下午第一份电报到达时间的分布是什么？

4．$\{N(t),t\geqslant 0\}$ 为一 $\lambda=2$ 的 Poisson 过程试求

(i) $P\{N(1)\leqslant 2\}$，

(ii) $P\{N(1)=1 \text{ 且 } N(2)=3\}$，

(iii) $P\{N(1)\geqslant 2|N(1)\geqslant 1\}$。

5．证明概率 $P_m(t)=P\{N(t)=m\}$ 在命题 2.1 的假定 (1)—(4) 下满足微分方程

$$P'_m(t)=-\lambda P_m(t)+\lambda P_{m-1}(t), \quad m=1,2,\cdots$$

并证明在初始条件 $P_m(0)=0$，$m=1,2,\cdots$ 下的解为 $\frac{\lambda^m t^m}{m!}e^{-\lambda t}$。

6．一部600页的著作总计有240个印刷错误，试利用Poisson近似求出某连续三页无错误的概率。

7．$N(t)$ 是强度为 λ 的 Poisson 过程。给定 $N(t)=n$，试求第 r 个事件 $(r\leqslant n)$ 发生的时刻 W_r 的条件概率密度 $f_{W_r|N(t)=n}(w_r|n)$。

8．令 $\{N_i(t), t\geqslant 0\}_{i=1}^{n}$ 为 n 个独立的有相同强度参数 λ 的 Poisson 过程。记 T 为在全部 n 个过程中至少发生了一件事的时刻。试求 T 的分布。

9．考虑参数为 λ 的 Poisson 过程 $N(t)$，若每一事件独立地以概率 p 被观察到。并将观察到的过程记为 $N_1(t)$，试问 $N_1(t)$ 是什么过程？$N(t)-N_1(t)$ 呢？ $N_1(t)$ 与 $N(t)-N_1(t)$ 是否独立？

10．到达某加油站的公路上的卡车服从参数为λ_1的Poisson过程 $N_1(t)$，而到达的小汽车服从参数为 λ_2 的Poisson过程$N_2(t)$。且过程 N_1 与N_2 独立。试问随机过程$N(t)=N_1(t)+N_2(t)$是什么过程？并计算在总车流数 $N(t)$ 中卡车首先到达的概率。

11．冲击模型(Shock Model)记$N(t)$为某系统到时刻 t 受到的冲击次数，它为参数为 λ 的 Poisson 过程。设第 k 次冲击对系统的损害的大小 Y_k 服从参数为 μ 的指数分布。Y_k，$k=1,2,\cdots$独立同分布。记 $X(t)$ 为系统所受到的总损害。当损害超过一定极限 a 时。系统不能运行，寿命终止。记 T 为系统寿命。试求该系统的平均寿命 ET，并对所得结果作出直观解释（提示：对非负随机变量 $ET=\int_0^\infty P(T>t)\mathrm{d}t$）。

12．令 $N(t)$ 是强度函数为 $\lambda(t)$ 的非齐次 Poisson 过程，$X_1, X_2, \cdots$ 为事件之间的间隔时间。

(i) X_i 是否独立，

(ii) X_i 是否同分布，

(iii) 试求 X_1 及 X_2 的分布。

13．考虑对所有 t，强度函数 $\lambda(t)$ 均大于0的非齐次Poisson过程 $\{N(t), t\geqslant 0\}$。令 $m(t)=\int_0^t \lambda(u)\mathrm{d}u$，$m(t)$ 的反函数为 $l(t)$。记 $N_1(t)=N(l(t))$ 试证 $N_1(t)$ 是通常的 Poisson 过程。试求 $N_1(t)$ 的强度参数 λ。

14．设 $N(t)$ 为更新过程，试判断下述命题的真伪：

(i) $\{N(t)<k\}\Longleftrightarrow\{W_k>t\}$，

(ii) $\{N(t)\leqslant k\}\Longleftrightarrow\{W_k\geqslant t\}$，

(iii) $\{N(t)>k\}\Longleftrightarrow\{W_k<t\}$，

其中 W_k 为第 k 个事件的等待时间。

第三章　Markov 过程

独立随机试验模型最直接的推广就是 Markov 链模型。因早在 1906 年就对它进行研究的俄国数学家 Markov 而得名。以后 Kolmogorov，Feller 和 Doob 等数学家发展了这一理论。粗略而言，一个随机过程如果给定了当前时刻 t 的值 X_t，未来 X_s $(s>t)$ 的值不受过去的值 X_u $(u<t)$ 的影响就称为是有 Markov 性。当指标集 T 是非负整数值时，过程称为是离散时间 Markov 链。而当 t 是连续时间时就是连续时间 Markov 链。进一步若状态空间也是连续的就是 Markov 过程。本章主要讨论 Markov 链。

3.1　Markov 链的定义和例子

从本节起我们主要考虑随机过程 $\{X_n, n=0, 1, \cdots\}$。下标 n 常代表时间。X_n 取值的状态空间为有限集或可列集。状态空间中的元素可以记为非负整数 $0, 1, 2, \cdots$。它们称作是随机过程所处的状态。比如，当 $X_n=i$，就称过程在时间 n 处于状态 i。

定义 3.1　如果对任何一列状态 $i_0, i_1, \cdots, i_{n-1}, i, j$，及对任何 $n\geqslant 0$，随机过程 $\{X_n, n\geqslant 0\}$ 满足 Markov 性质

$$P\{X_{n+1}=j|X_0=i_0, \cdots, X_{n-1}=i_{n-1}, X_n=i\}$$
$$=P\{X_{n+1}=j|X_n=i\} \qquad (3.1)$$

则称 X_n 为离散时间 Markov 链。

定义 3.2　设 X_n 为一离散时间 Markov 链。给定 X_n 在状态 i 时 X_{n+1} 处于状态 j 的条件概率 $P\{X_{n+1}=j|X_n=i\}$ 称作是 Markov 链的一步转移概率，记作 $P_{ij}^{n,n+1}$。当这一概率与 n 无关时称该 Markov 链有平稳转移概率，并记之为 P_{ij}。

我们将主要讨论带平稳转移概率的Markov 链，它也称作时间齐性 Markov 链或简称时齐 Markov 链。由于概率是非负的，而且过程总要转移到某一状态去所以很自然地有：对任何 i，$j\geq 0$，

$$P_{ij}\geq 0 \text{ 且 } \sum_{j=0}^{\infty} P_{ij}=1. \tag{3.2}$$

当然在此我们把过程留在原地也看成是一种"转移"，即从 i 转移到 i。通常把 P_{ij} 排成一个无穷维的方阵。记作 $\boldsymbol{P}$

$$\boldsymbol{P}=\begin{Vmatrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ & \vdots & & \\ P_{i0} & P_{i1} & P_{i2} & \cdots \\ & \vdots & & \end{Vmatrix}. \tag{3.3}$$

$\boldsymbol{P}=\|P_{ij}\|$ 称为 Markov 链的转移概率矩阵。矩阵的第 $i+1$ 行就是给定 $X_n=i$ 时，X_{n+1} 的条件概率分布。当 Markov 链的状态总数是有限的，则 $\boldsymbol{P}$ 就是有限阶的方阵。其阶数正好是状态空间中状态的总数。先看一个简单的例子：

例 3.1 直线上的随机游动。考虑在直线上整数点上运动的粒子。当它处于位置 j 时，这里姑且假定 j 就是过程所处的状态，向右游动到 $j+1$ 的概率为 p 而向左游动到 $j-1$ 的概率为 $q=1-p$。假定时刻 0 时粒子处在原点，即 $X_0\equiv 0$。于是粒子在时刻 n 所处的位置 X_n 就是一个 Markov 链。它有转移概率

$$P_{jk}=\begin{cases} p, & k=j+1, \\ q, & k=j-1, \\ 0, & \text{其他 } k, \end{cases}$$

当 $p=q=\frac{1}{2}$ 时，就是简单对称随机游动。它可用在公平赌博的模型中。比如考虑出现正、反面概率均为 $\frac{1}{2}$ 的扔币游戏。让粒子的位置 X_n 代表赌徒甲在第 n 次博戏之后所赢的钱数。（每扔币一次的输赢为 1 元）假如甲方有赌本 A_1，乙方有赌本 A_2。则甲所赢

感兴趣的问题是在自己输光之前乙先输光的可能性有多大。这就是著名的赌徒输光问题。可以证明这一概率为 $A_1/(A_1+A_2)$。赌本越多越容易赢得最后胜利。这一模型在统计学的序贯分析中有重要应用。状态 A_2 及 $-A_1$ 称为吸收态。这样的随机游动又叫带吸收壁的随机游动。

一个 Markov 链可由它的初始状态 X_0，或初始状态的概率分布以及转移概率阵所完全确定。记 $P\{X_0=i\}=p_i$，不难算出：

$$
\begin{aligned}
&P\{X_0=i_0,X_1=i_1,\cdots,X_n=i_n\}\\
&=P\{X_0=i_0,X_1=i_1,\cdots,X_{n-1}=i_{n-1}\}\\
&\quad P\{X_n=i_n|X_0=i_0,X_1=i_1,\cdots,X_{n-1}=i_{n-1}\}\\
&=P\{X_0=i_0,X_1=i_1,\cdots,X_{n-1}=i_{n-1}\}P\{X_n=i_n|X_{n-1}=i_{n-1}\}\\
&=P\{X_0=i_0,X_1=i_1,\cdots,X_{n-1}=i_{n-1}\}P_{i_{n-1},i_n}.
\end{aligned}
$$

由归纳法可知

$$
\begin{aligned}
&P\{X_0=i_0,X_1=i_1,\cdots,X_n=i_n\}\\
&=p_{i_0}P_{i_0,i_1}P_{i_1,i_2}\cdots P_{i_{n-2},i_{n-1}}P_{i_{n-1},i_n}
\end{aligned}
\tag{3.4}
$$

于是，就可以掌握随机过程 $\{X_n,n\geqslant 0\}$ 在各不同时刻的运行规律。在 Markov 链的分析中，人们主要关心过程的可能实现的概率有多大。在其计算中，n 步转移概率 $P_{ij}^{(n)}=P\{X_{m+n}=j|X_m=i\}$ 是一个关键的量，它表示给定当时刻 m 时过程处于状态 i 间隔 n 步之后过程在时刻 $m+n$ 时转移到了状态 j 的条件概率。以 $P_{ij}^{(n)}$ 为第 i，j 元的矩阵 $\|P_{ij}^{(n)}\|$ 记作 $\boldsymbol{P}^{(n)}$ 是 Markov 链的 n 步转移概率矩阵。有趣的是 $\boldsymbol{P}^{(n)}$ 可以通过转移概率矩阵 $\boldsymbol{P}$ 求出，确切地讲，我们有 $\boldsymbol{P}^{(n)}=\boldsymbol{P}\times\boldsymbol{P}\cdots\times\boldsymbol{P}=\boldsymbol{P}^n$。也即 $\boldsymbol{P}^{(n)}$ 可由一步转移概率矩阵 $\boldsymbol{P}$ 自乘 n 次得到。这就是

定理 3.1 Markov 链的 n 步转移概率矩阵满足

$$
P_{ij}^{(n)}=\sum_{k=0}^{\infty}P_{ik}P_{kj}^{(n-1)}. \tag{3.5}
$$

在上式中我们约定 $P_{ii}^{(0)}=1$，当 $j\neq i$ 时 $P_{ij}^{(0)}=0$。

证明 思路是先分析转移一步的所有可能情形，并充分利用

过程的 Markov 性质. 假如过程从状态 i 出发 n 步之后转移到了状态 j，那么在第一步之后必然要跳到某一状态 k 然后再从状态 k 经过 $n-1$ 步转移到达状态 j. 注意到第一步转移到不同状态 k 的那些事件是互不相容的.这好比是一个人不能同时骑在两匹马上. 于是有：

$$\begin{aligned}P_{ij}^{(n)} &= P\{X_n=j|X_0=i\}\\ &=\sum_{k=0}^{\infty} P\{X_n=j, X_1=k|X_0=i\}\\ &=\sum_{k=0}^{\infty} P\{X_1=k|X_0=i\}P\{X_n=j|X_0=i, X_1=k\}\\ &=\sum_{k=0}^{\infty} P_{ik}P_{kj}^{(n-1)}.\end{aligned}$$

若将(3.5)式写成矩阵形式，由矩阵的乘法规则正好是 $\boldsymbol{P}^{(n)}=\boldsymbol{P}\times\boldsymbol{P}^{(n-1)}$. 于是有 $\boldsymbol{P}^{(n)}=\boldsymbol{P}\times\boldsymbol{P}^{n-1}=\boldsymbol{P}^n$，更一般的，对任何 $m, n\geqslant 0$，$\boldsymbol{P}^{(m+n)}=\boldsymbol{P}^{(n)}\times\boldsymbol{P}^{(m)}$. 把矩阵相乘形式写成其元素的方程就是著名的 Chapman-Kolmogorov 方程：

$$P_{ij}^{(n+m)}=\sum_{k=0}^{\infty} P_{ik}^{(n)}P_{kj}^{(m)}. \tag{3.6}$$

下面介绍一些 Markov 链的例子：

例 3.2 Ehrenfest 模型. 在描述薄膜间分子扩散时会遇到 Ehrenfest 的罐子模型. 此时，设想薄膜的两侧为两个不同的罐子，而分子就是装在罐子中的小球. 假定两个罐子总共装有 $2a$ 个球(即分子总数). A 罐中有 k 个球，B 罐中有 $2a-k$ 个球. 从全部 $2a$ 个球中每次等概率地任选一球放进另一个罐子就形象地描述了分子在薄膜间的随机扩散. 显然球要在两罐之间来回运动，而且总是倾向于从较满的罐子挪向较空的罐子. 因为罐中的球越多被抽中的概率也越大. 记在时刻 n 时 A 罐中有 Y_n 个球，则 B 罐中有 $2a-Y_n$ 个球. 定义 A 罐中的球数超出总数一半的量为 X_n 即 $X_n=Y_n-a$. 则 $\{X_n\}$ 为一 Markov 链. 它可取有限个状态 $i=-a, -a+1, \cdots, 0, 1, \cdots, a$. 而转移概率为

$$P_{ij}=\begin{cases}\dfrac{a-i}{2a}, & \text{当 } j=i+1,\\ \dfrac{a+i}{2a}, & \text{当 } j=i-1,\\ 0 & \text{，其他情形。}\end{cases}$$

在这一模型中人们感兴趣的是球在每个罐子中的平衡分布。

例 3.3 理想化的遗传模型. 遗传的要素是染色体。遗传性质的携带者称为基因，它位于染色体上。基因控制着生物的特征。基因是成对出现的。控制同一特征的不同基因称为是等位基因。记这对等位基因为 A 和 a。在一个总体中基因 A 和 a 的比例是基因频率，常记为 p 和 q.($p+q=1$)。现固定总体中的个体数为 $2N$。则繁衍出的第二代的基因型是由 $2N$ 次 Bernoulli 试验所确定的。如第 n 代母体中 A 型基因出现 j 次，a 型基因出现 $2N-j$ 次。则下一代为 A 型的概率为 $p_j=\dfrac{j}{2N}$，为 a 型的概率为 $q_j=1-\dfrac{j}{2N}$。记 X_n 为在第 n 代中携带 A 型基因的个体数，则 $\{X_n, n\geqslant 1\}$ 为一 Markov 链.其转移概率矩阵的第 j, k 元 P_{jk}

$$P_{jk}=\binom{2N}{k}p_j^k q_j^{2N-k}=\binom{2N}{k}\left(\frac{j}{2N}\right)^k\left(1-\frac{j}{2N}\right)^{2N-k}.$$

著名生物统计学家 Fisher 曾对这一模型作过深入的讨论。若把变异考虑进去就可以建立更复杂的遗传模型。

例 3.4 排队问题. 顾客到服务台排队等候服务。在每一服务周期中只要服务台前有顾客在等待就要对排在队前的一位提供服务。当然，如果服务台前是空的就不可能实施服务。设在第 n 个服务周期中到达的顾客数为一随机变量 Y_n，其分布不依赖于 n 为

$$P\{Y_n=k\}=p_k,\ k=0,1,\cdots,\sum_{k=0}^{\infty}p_k=1,$$

并且 Y_n 是相互独立的，记 X_n 为服务周期 n 开始时服务台前的顾

客数。则显然有

$$X_{n+1}=\begin{cases}X_{n-1}+Y_n, & 若\ X_n\geqslant 1,\\ Y_n, & 若\ X_n=0.\end{cases}$$

此时 $\{X_n, n\geqslant 1\}$ 为一 Markov 链。其转移概率阵为

$$\boldsymbol{P}=\begin{Vmatrix} p_0 & p_1 & p_2 & p_3 & p_4 & \cdots \\ p_0 & p_1 & p_2 & p_3 & p_4 & \cdots \\ 0 & p_0 & p_1 & p_2 & p_3 & \cdots \\ 0 & 0 & p_0 & p_1 & p_2 & \cdots \\ 0 & 0 & 0 & p_0 & p_1 & \cdots \end{Vmatrix}.$$

可以想见当新到顾客数的期望 $\sum\limits_{k=0}^{\infty} kp_k>1$ 时排队长度将趋于无穷大，反之当 $\sum\limits_{k=0}^{\infty} kp_k<1$ 时就有可能趋于一个平衡状态。此时的极限概率

$$\lim_{n\to\infty}P\{X_n=k\,|\,X_0=j\}=\pi_k.\quad k=0, 1, \cdots,$$

称为是排队模型的平衡分布。π_0 表示服务设施闲置的概率。这是管理科学中衡量服务系统效率的一个重要指标。而在另一方面顾客在服务台前化去的平均时间 $\sum\limits_{k=0}^{\infty}(1+k)\pi_k$ 是衡量顾客满意程度和服务质量的重要指标。有关平衡分布的问题我们在后面还会加以讨论。

例 3.5 考虑一个三状态的 Markov 链 $\{X_n\}$。其转移概率矩阵为

$$\boldsymbol{P}=\begin{matrix}0\\1\\2\end{matrix}\begin{Vmatrix}1 & 0 & 0\\ p & q & r\\ 0 & 0 & 1\end{Vmatrix}.$$

其中 $p, q, r>0, p+q+r=1$。这一 Markov 链从状态 1 开始一旦进入状态 0 或 2 就无法跳出而被吸收了。于是可以提出两个问题：假如过程从状态 1 出发被状态 0（或 2）吸收的概率是多少？平均要多长时间过程会进入吸收态而永远停在哪儿？

解 让我们先把过程进入吸收态的时刻表示出来。令T =｛过程进入吸收态的时刻｝$=\min\{n\geqslant 0 | X_n=0$ 或$X_n=2\}$。则例中所求概率及期望可以表示为$u=$｛从状态1出发最终进入吸收态0的概率｝$=P\{X_T=0|X_0=1\}$ 及$v=$｛从状态1出发过程进入吸收态的平均时间｝$=E\{T|X_0=1\}$。注意到如$X_1=0$，则$T=1$，于是$X_T=0$，而如果$X_1=2$也有$T=1$但$X_T=2$；只有当$X_1=1$时过程才回到X_0所处的状态1并重新开始转移。因此由全概率公式我们有

$$
\begin{aligned}
u&=P\{X_T=0|X_0=1\}\\
&=\sum_{k=0}^{2}P\{X_T=0|X_0=1, X_1=k\}P\{X_1=k|X_0=1\}\\
&=\sum_{k=0}^{2}P\{X_T=0|X_1=k\}P\{X_1=k|X_0=1\}\\
&=1\cdot p+u\cdot q+0\cdot r=p+qu.
\end{aligned}
$$

于是解出 $u=\dfrac{p}{1-q}=\dfrac{p}{p+r}$。类似地可以建立关于$v$的方程求出最终被吸收的平均时间：

$$
\begin{aligned}
v&=E\{T|X_0=1\}\\
&=\sum_{k=0}^{2}E\{T|X_0=1, X_1=k\}P\{X_1=k|X_0=1\}\\
&=\sum_{k=0}^{2}E\{T|X_1=k\}P\{X_1=k|X_0=1\}\\
&=1\cdot p+(1+v)\cdot q+1\cdot r\\
&=1+qv.
\end{aligned}
$$

解出 $v=\dfrac{1}{1-q}$。可以从解中看到过程从状态1转回状态1的概率q越大就越难进入吸收状态，而且平均时间拉长。这是符合常识的。

3.2 Markov 链的状态分类

3.2.1 互达性和周期性

为了更深入地研究 Markov 链需要对状态进行分类。这将涉及一连串的定义。重要的是要把握它们的概率含义。并通过实际例子来加深对这些概念的理解。

定义 3.2 可达与互达。如果对某一 $n\geqslant 0$，有 $P_{ij}^{(n)}>0$ 则称状态 j 是从状态 i 可达的 (accessible)记作 $i\longrightarrow j$，它表示从状态 i 经过有限步的转移可以到达状态 j。两个互相可达的状态 i 和 j 则称为是互达的(communicate) 记作 $i\longleftrightarrow j$。

如果两个状态 i 和 j 不是互达的，那就有对所有 $n\geqslant 0$，$P_{ij}^{(n)}=0$ 或者对所有 $n\geqslant 0$ $P_{ji}^{(n)}=0$，或者两者都成立。三种情况必居其一。互达性是一种数学上的等价关系，也就是说它满足自反性，对称性、和传递性。

命题 3.1 互达性是等价关系

i) $i\longleftrightarrow j$，自反性，

ii) 若 $i\longleftrightarrow j$，则 $j\longleftrightarrow i$，对称性，

iii) 若 $i\longleftrightarrow j$，且 $j\longleftrightarrow k$，则 $i\longleftrightarrow k$，传递性。

证明 i) 和 ii) 由互达性的定义马上即可证明。现仅证明 iii)。假如 $i\longleftrightarrow j$ 则存在整数 n 使 $P_{ij}^{(n)}>0$，又由 $j\longleftrightarrow k$ 知存在整数 m 使 $P_{jk}^{(m)}>0$。由(3.6)式 Chapman-Kolmogorov 方程

$$P_{ik}^{(n+m)}=\sum_{r=0}^{\infty}P_{ir}^{(n)}P_{rk}^{(m)}\geqslant P_{ij}^{(n)}P_{jk}^{(m)}>0.$$

类似地可以证明存在 s 使 $P_{ki}^{(s)}>0$。于是由定义 $i\longleftrightarrow k$。

两个状态如果是互达的就称他们是处在同一类中。Markov链的所有状态就由互达这一等价关系而分割成不同的等价类。由命题 3.1 我们立刻知道两个类要么互不相交，要么完全重合。如果

在互达性这一等价关系下 Markov 链的所有状态都居于同一类那么就称这个 Markov 链是不可约的(irreducible)。换言之，不可约过程的各个状态都是互达的。

例 3.5 若 Markov 链有转移概率矩阵

$$P=\begin{pmatrix} \frac{1}{4} & \frac{3}{4} & & & \\ \frac{1}{2} & \frac{1}{2} & & 0 & \\ & & 0 & 1 & 0 \\ & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ & & 0 & 1 & 0 \end{pmatrix}$$

则显见 $\{1,2\}$ 和 $\{3,4,5\}$ 是状态在互达意义下的两个等价类。这个链是可约的。我们可以把它分成两个链来研究。

例 3.1 续 在随机游动的例子 3.1中状态$\{A_2\}$，$\{-A_1+1,\cdots,0,1,2,\cdots,A_2-1\}$，$\{-A_1\}$是三个不同的类。这时可以从中间那一类转移到吸收态 A_2 或 $-A_1$。但一旦被吸收后就永远无法返回。所以过程并不是不可约的。

定义 3.3 状态 i 的周期。i 为 Markov 链的一个状态，使 $P_{ii}^{(n)}>0$ 的所有 n $(n\geqslant 1)$ 的最大公约数称作是状态 i 的周期记作 $d(i)$。如果对所有 $n\geqslant 1$，都有 $P_{ii}^{(n)}=0$ 则约定周期为 ∞；$d(i)=1$ 的状态 i 则称为是非周期的。

由定义立即可知如 n 不能被周期 $d(i)$ 整除则必有 $P_{ii}^{(n)}=0$。

例 3.6 Markov 链有状态 0，1，2，3 和转移概率阵

$$P=\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

试求状态 0 的周期。

解 不难直接算出 $P_{00}=0$，$P_{00}^{(2)}=P_{00}^{(3)}=P_{00}^{(5)}=P_{00}^{(2n+1)}=0$ 而 $P_{00}^{(4)}=\frac{1}{2}$，$P_{00}^{(6)}=\frac{1}{4}$，$P_{00}^{(8)}=\frac{3}{8}$。而 $\{4,6,8,\cdots\}$ 的最大公约数为 2。所以 $d(0)=2$。

那么上例中其他状态的周期是多少呢？由于周期性是一种整个等价类所具有的性质所以如果过程是不可约的（如例 3.6）则每个状态都有相同的周期。这就是

命题 3.2 如果 $i\longleftrightarrow j$ 则 $d(i)=d(j)$

证明 由 $i\longleftrightarrow j$ 知存在 m，n 使 $P_{ij}^{(m)}>0$ 和 $P_{ji}^{(n)}>0$。于是有

$$P_{jj}^{(n+m)}\geqslant P_{ji}^{(n)}P_{ij}^{(m)}>0.$$

假如有 s 使 $P_{ii}^{(s)}>0$ 则也有

$$P_{jj}^{(n+s+m)}\geqslant P_{ji}^{(n)}P_{ii}^{(s)}P_{ij}^{(m)}>0.$$

因为不等式最左边所表示的从状态 j 出发经过 $n+s+m$ 步转移后又回到 j 的概率它当然要大于一个加了更多限制的子事件的概率。这个子事件是从 j 出发经过 n 步转移到 i，再经过 s 步返回 i，又再从 i 出发经过 m 步到达 j。它的效果也是转移 $n+s+m$ 步回到 j。由 $d(j)$ 的定义它将同时整除 $n+m$ 及 $n+s+m$ 所以 $d(j)$ 必整除 s，而 $d(i)$ 是所有使 $P_{ii}^{(s)}>0$ 的 s 的最大公约数所以 $d(j)$ 整除 $d(i)$。同样可证 $d(i)$ 整除 $d(j)$，所以有 $d(i)=d(j)$。

下面讨论周期的基本性质。我们有

命题 3.3 如果状态 i 有周期 $d(i)$，则存在整数 N 使得对所有的 $n>N$ 恒有 $P_{ii}^{(nd(i))}>0$。

证明 命题的证明主要依赖一个数论上的事实：如果正整数 n_1，n_2，$\cdots$，n_k 的最大公约数为 d 则存在正整数 N 使得对所有的 $n>N$ 我们能找到非负整数 c_i 使得

$$nd=\sum_{i=1}^{k}c_in_i, \tag{3.7}$$

对状态 i，令 n_1，n_2，$\cdots$，n_k 为使 $P_{ii}^{(n_r)}>0$，$r=1,2,\cdots,k$ 的正整

数。由周期的定义，$d(i)$是它们的最大公约数。而由（3.7）式知对 $n>N$

$$P_{ii}^{(nd(i))}=P_{ii}^{(\sum_{r=1}^{k}c_r n_r)}\geqslant\prod_{r=1}^{k}(P_{ii}^{(n_r)})^{c_r}>0$$

由命题 3.3 及事实 $P_{ji}^{(m+d(i))}\geqslant P_{ji}^{(m)}P_{ii}^{(nd(i))}$ 立即有

推论 3.1 如果 $P_{ji}^{(m)}>0$，则存在正整数 N 使得对 $n\geqslant N$ 恒有 $P_{ji}^{(m+nd(i))}>0$。讨论这些性质的意义何在呢?它可以帮助我们研究当 n 很大时 $P_{ij}^{(n)}$ 的极限是否存在?存在的条件是什么?如果存在又如何简便地求出它们？周期性是一种等价类中全体状态共有的性质。当周期为 1 时 Markov 链称为是非周期的。对非周期的不可约的 Markov 链我们有

命题 3.4 令 $\boldsymbol{P}$ 为不可约、非周期、有限状态 Markov 链的转移概率矩阵。则必存在 N 使得当 $n\geqslant N$ 时 n 步转移概率阵 $\boldsymbol{P}^{(n)}$ 的所有元素都非零。

证明 由于 Markov 链是不可约的。过程的任两个状态 i 和 j 都是可达的。于是存在m(与 i, j 有关) 使 $P_{ij}^{(m)}>0$。由推论 3.1 及链是非周期的知存在 N 使得当 $n\geqslant N$ 时有 $P_{ij}^{(m+n\cdot 1)}>0$ 因状态空间有限，对全部的状态对 (i,j) 求出 $N(i,j)$。并取 $N=\max\limits_{(i,j)}(m(i,j)+N(i,j))$则显然对所有状态 i, j 有当 $n>N$ 时 $P_{ij}^{(n)}>0$。

这种存在n 使 $\boldsymbol{P}^n$ 的元素全部非零的 k 状态（记为 $0,1,\cdots,k-1$）Markov 链称为是正则的。而最重要的事实是一个正则的有限状态 Markov 链极限 $\lim\limits_{n\to\infty}P_{ij}^{(n)}$.总是存在的，记之为 π_j。并且极限为一概率分布。即有 $\pi=(\pi_0,\pi_1,\cdots,\pi_{k-1})$存在 $\pi_j>0$, $j=0,1,\cdots,k-1$ 及 $\sum_j\pi_j=1$。这些是下一节极限定理和平稳分布所要讨论的内容。

3.2.2 常返(recurrent) 与瞬过 (transient)

为了以下讨论的方便，我们引入一个重要的概率 $f_{ij}^{(n)}$，它代表

从 i 出发在 n 步转移时首次到达 j 的概率。用式子表达即是

$$f_{ij}^{(0)}=0,$$

$$f_{ij}^{(n)}=P\{X_n=j, X_k\neq j, k=1, \cdots, n-1 \mid X_0=i\}$$

而 $f_{ii}^{(n)}$ 就代表从 i 出发在第 n 次转移时首次回到状态 i 的概率。因不曾出发就不可能有回来所以约定 $f_{ii}^{(0)}=0$。同样有

$$f_{ii}^{(n)}=P\{X_n=i, X_k\neq i, k=1, \cdots, n-1 \mid X_0=i\},$$

记 $f_{ij}=\sum_{n=1}^{\infty}f_{ij}^{(n)}$，它是从 i 出发最终转入状态 j 的概率。当 $i\neq j$，则 $i\longleftrightarrow j$ 当且仅当 $f_{ij}>0$。现在可以给出

定义 3.4 如果 $f_{ii}=1$ 我们称状态 i 是常返的，一个非常返状态就称为是瞬过的。

从定义知道如果 i 是常返的状态，那么从 i 出发经过有限步转移后最终又回到 i 的概率为 1。那么如何判断一个状态 i 是否是常返的呢？我们给出用 n 步转移概率 $P_{ii}^{(n)}$ 表示的一个判别准则：

定理 3.2 状态 i 常返的充分必要条件是

$$\sum_{i=1}^{\infty}P_{ii}^{(n)}=\infty. \tag{3.8}$$

当然与此等价地有状态 i 是瞬过的当且仅当

$$\sum_{i=1}^{\infty}P_{ii}^{(n)}<\infty. \tag{3.9}$$

证明 先考虑瞬过状态 i，这时过程从 i 出发至少返回 i 一次的概率为 f_{ii} 且 $f_{ii}<1$。由过程的 Markov 性一旦回到 i 过程以后的发展只依赖当前，因此从 i 出发至少回到 i 两次的概率为 f_{ii}^2。这样不断地重复，并令随机变量 K 为过程返回 i 的次数。则 K 在给定 $X_0=i$ 下的条件概率分布应满足

$$P\{K\geqslant k \mid X_0=i\}=f_{ii}^k, k=1, 2, \cdots$$

这正是参数为 f_{ii} 的几何分布。于是 K 的条件期望为：

$$E(K \mid X_0=i)=\sum_{k=1}^{\infty}P(K\geqslant k \mid X_0=i)$$

$$=\sum_{k=1}^{\infty} f_{ii}^{k}=\frac{f_{ii}}{1-f_{ii}}.$$

显然，当 $f_{ii}<1$ 时 $E(K|X_0=i)<\infty$。即过程从 i 出发平均返回 i 的次数是有限的。另一方面记

$$I_n=\begin{cases}1, & \text{若 } X_n=i,\\ 0, & \text{其他},\end{cases}$$

则有 $K=\sum_{n=1}^{\infty} I_n$。于是，

$$\begin{aligned}\sum_{n=1}^{\infty} P_{ii}^{(n)}&=\sum_{n=1}^{\infty} P\{X_n=i|X_0=i\}\\ &=\sum_{n=1}^{\infty} E\{I_n|X_0=i\}=E\left\{\sum_{n=1}^{\infty} I_n|X_0=i\right\}\\ &=E\{K|X_0=i\}<\infty.\end{aligned}$$

反之，如果 $\sum_{n=1}^{\infty} P_{ii}^{(n)}<\infty$ 则必有 i 为瞬过的。如若不然，i 是常返的则从 i 出发最终回到 i 的概率为 1。由Markov性过程继续不断地重复，不断地返回 i。因此以概率为 1 地有过程返回 i 的次数是无穷大，即 $E\{K|X_0=i\}=\infty$ 与 $\sum_{n=1}^{\infty} P_{ii}^{(n)}$ 收敛相矛盾。定理证毕。

推论 3.2 如果 i 是常返的，且 $i\longleftrightarrow j$ 则 j 也是常返的。

证明 由 $i\longleftrightarrow j$ 知存在 m, n 使 $P_{ji}^{(m)}>0$ 和 $P_{ij}^{(n)}>0$。于是对任何 $s>0$ 有

$$P_{jj}^{(m+s+n)}\geqslant P_{ji}^{(m)} P_{ii}^{(s)} P_{ij}^{(n)},$$

所以

$$\begin{aligned}\sum_{s=1}^{\infty} P_{jj}^{(m+s+n)}&\geqslant\sum_{s=1}^{\infty} P_{ji}^{(m)} P_{ii}^{(s)} P_{ij}^{(n)}\\ &=P_{ji}^{(m)} P_{ij}^{(n)}\sum_{s=1}^{\infty} P_{ii}^{(s)}=\infty.\end{aligned}$$

因此 $\sum_{k=1}^{\infty} P_{jj}^{(k)}=\infty$。由定理 3.2 知 j 为常返状态。常返性也是一种等价类共有的性质。在同一个等价类中的状态要么都是常返

的，要么都是瞬过的。

例 3.7 继续考虑整数点上的随机游动。向右挪一格的概率为 p，向左挪一格的概率为 $q=1-p$。从原点（状态 0）出发只有转移偶数步才能回到原点。所以 $P_{00}^{(2n+1)}=0$，$n=1,2,\cdots$ 而 $P_{00}^{(2n)}=\binom{2n}{n}p^n q^n$ 利用 Stirling 公式知当 n 充分大时 $n!\sim n^{n+\frac{1}{2}}e^{-n}\sqrt{2\pi}$

于是 $P_{00}^{(2n)}\sim\frac{2^{2n}(pq)^n}{\sqrt{\pi n}}=\frac{(4pq)^n}{\sqrt{\pi n}}$。而 $pq\leqslant\frac{1}{4}$，等号成立当且仅当 $p=q=\frac{1}{2}$。于是有当 $p=\frac{1}{2}$ 时 $\sum\limits_{n=1}^{\infty}P_{ii}^{(n)}=\infty$。因此直线上的对称随机游动是常返的。显然，当 $p\neq q$ 过程的状态是瞬过的。有趣的是可以证明二维对称随机游动也是常返的而三维以上的对称随机游动却是瞬过的。详细证明请参看习题 3.13。

对常返状态 i 我们定义 T_i 为首次返回状态 i 的时刻，称作常返时。记 $\mu_i=ET_i$ 则有

$$\mu_i=\sum_{n=1}^{\infty}nf_{ii}^{(n)}. \tag{3.10}$$

回想 $f_{ii}^{(n)}$ 代表从 i 出发在第 n 步转移时首次回到 i 的概率。所以 μ_i 是首次返回 i 的期望步数，也叫状态 i 的平均常返时。利用 μ_i 可以对常返状态作进一步的分类：分为零常返和正常返。

定义 3.5 一个常返状态 i 当且仅当 $\mu_i=\infty$ 时称为是零常返的。而当且仅当 $\mu_i<\infty$ 时称为正常返的。

对于有限多个状态的 Markov 链 μ_i 总是有限的，所以只有在有可列无穷多个状态时才可能是出现零常返的状态（参看习题 3.15）。当状态数目不大时直接的计算分析就可以把状态分类弄清、但当状态总数很大时就要借助计算机、已有学者研究了利用图论的算法对 Markov 链进行分类。有兴趣的读者可参看有关的著作。

3.3 Markov 链的极限定理与平稳分布

在上一节有关 Markov 链状态分类的讨论中已经引进了常返状态 i 的常返时 T_i。它还可记为

$$T_i=\min\{n\geqslant 1: X_n=i\},$$

而
$$\begin{aligned} f_{ii}^{(n)}&=P\{X_n=i, X_k\neq i, k=1,\cdots,n-1\mid X_0=i\}\\ &=P\{T_i=n\mid X_0=i\},\quad n=1,2,\cdots, \end{aligned}$$

表示了 T_i 的条件概率分布，$\mu_i=\sum_{n=1}^{\infty} nf_{ii}^{(n)}$ 则为 T_i 在给定 $X_0=i$ 时的条件期望。在进行更深入的讨论之前先看

例 3.8 考虑只有 0,1 两个状态的 Markov 链 X_n，转移概率矩阵为

$$P=\begin{Vmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{Vmatrix},\quad 0<\alpha,\ \beta<1$$

当 $\alpha=1-\beta$ 时 X_n 为独立同分布的随机变量．且 $P(X_n=0)=\beta$，$P(X_n=1)=\alpha=1-\beta$。而当 $\alpha\neq 1-\beta$ 时 X_n 是通常意义下的 Markov 链。一方面不难算出对状态 0，T_0 有分布

$$f_{00}^{(1)}=(1-\alpha),\ f_{00}^{(n)}=\alpha\beta(1-\beta)^{n-2},\ n=2,3,\cdots,$$

于是 $\mu_0=ET_0=\sum_{n=1}^{\infty} nf_{00}^{(n)}=\dfrac{\alpha+\beta}{\beta}$。另一方面，记

$$A=\begin{Vmatrix} \beta & \alpha \\ \beta & \alpha \end{Vmatrix},\quad B=\begin{Vmatrix} \alpha & -\alpha \\ -\beta & \beta \end{Vmatrix},$$

则 $P=\dfrac{1}{\alpha+\beta}A+\dfrac{1-\alpha-\beta}{\alpha+\beta}B$。而且由归纳法可以证明

$$P^n=\frac{1}{\alpha+\beta}\begin{Vmatrix} \beta & \alpha \\ \beta & \alpha \end{Vmatrix}+\frac{(1-\alpha-\beta)^n}{\alpha+\beta}\begin{Vmatrix} \alpha & -\alpha \\ -\beta & \beta \end{Vmatrix}$$

这是因为

$$\begin{aligned}
\boldsymbol{P}^n\cdot\boldsymbol{P}&=\frac{1}{\alpha+\beta}(\boldsymbol{A}+(1-\alpha-\beta)^n\boldsymbol{B})\cdot\boldsymbol{P}\\
&=\frac{1}{\alpha+\beta}(\boldsymbol{A}\cdot\boldsymbol{P}+(1-\alpha-\beta)^n\boldsymbol{B}\cdot\boldsymbol{P})\\
&=\frac{1}{\alpha+\beta}(\boldsymbol{A}+(1-\alpha-\beta)^{n+1}\boldsymbol{B})=\boldsymbol{P}^{n+1}.
\end{aligned}$$

由 $0<\alpha,\ \beta<1$ 立即推出当 $n\to\infty$ 时，$|1-\alpha-\beta|^n\to 0$。于是有

$$\lim_{n\to\infty}\boldsymbol{P}^n=\begin{Vmatrix}\dfrac{\beta}{\alpha+\beta} & \dfrac{\alpha}{\alpha+\beta}\\[2mm] \dfrac{\beta}{\alpha+\beta} & \dfrac{\alpha}{\alpha+\beta}\end{Vmatrix}.$$

这里矩阵极限的含义是由它的每个元素序列极限所组成的矩阵。记 $\boldsymbol{Q}=\lim\limits_{n\to\infty}\boldsymbol{P}^{(n)}$ 则 $\boldsymbol{Q}=(q_{ij})$ 而 $q_{ij}=\lim\limits_{n\to\infty}P_{ij}^{(n)}$。这就告诉我们不论过程的初始状态是什么，过程在经过一段长时间后会以概率 $\dfrac{\beta}{\alpha+\beta}$ 处于状态 0，概率 $\dfrac{\alpha}{\alpha+\beta}$ 处于状态 1。我们把极限概率 $\{\lim\limits_{n\to\infty}P_{00}^{(n)},\lim\limits_{n\to\infty}P_{11}^{(n)}\}$ 记为 $\{\pi_0,\pi_1\}$。如果 $\pi=\{\pi_0,\ \pi_1\}$ 存在且为一概率分布就称为是例 3.8 这一 Markov 链的平稳分布。现在 $\pi_0=\dfrac{\beta}{\alpha+\beta}$ 细心的读者会发现 π_0 与 μ_0 互为倒数。这并不是偶然的巧合。下面的定理 3.3 讨论了其中深刻的背景。从直观上看 $\pi_0=\lim\limits_{n\to\infty}P_{00}^{(n)}=\dfrac{1}{\mu_0}$ 的意义是过程返回状态 0 的平均常返时越长则过程就长期而言处于状态 0 的概率越小。π_0 有两重含义，它既可以反映在时间长河中过程处于状态 0 的份额或机会又同时代表当整个过程处于平衡状态(极限情形下)时过程处于状态 0 的机会。

定理 3.3 Markov 链的基本极限定理

a) 若状态 i 是瞬过的或者是零常返的，则

$$\lim_{n\to\infty}P_{ii}^{(n)}=0. \tag{3.11}$$

b) 若状态 i 是周期为 d 的常返状态，则

$$\lim_{n\to\infty} P_{ii}^{(nd)} = \frac{d}{\mu_i}. \tag{3.12}$$

c) 当状态 i 是非周期的正常返状态，则

$$\lim_{n\to\infty} P_{ii}^{(n)} = \frac{1}{\mu_i}. \tag{3.13}$$

证明 我仅证明 c)，稍加修改就可以证明 a) 和 b)。记 $P_{ii}^{(n)}$，$f_{ii}^{(n)}$ $n=0, 1, 2, \cdots$ 的形式上的矩母函数分别为 $P_i(t)$ 和 $F_i(t)$，也即

$$P_i(t) = \sum_{n=0}^{\infty} e^{tn} P_{ii}^{(n)}, \quad F_i(t) = \sum_{n=0}^{\infty} e^{tn} f_{ii}^{(n)}, \quad t<0.$$

由习题 3.9 知

$$P_{ii}^{(n)} = \sum_{k=1}^{n} f_{ii}^{(k)} P_{ii}^{(n-k)}.$$

于是

$$\begin{aligned}
P_i(t) &= P_{ii}^{(0)} + \sum_{n=1}^{\infty} e^{tn} P_{ii}^{(n)} = 1 + \sum_{n=1}^{\infty} e^{tn} \sum_{k=1}^{n} f_{ii}^{(k)} P_{ii}^{(n-k)} \\
&= 1 + \sum_{k=1}^{\infty} f_{ii}^{(k)} \sum_{n=k}^{\infty} e^{tn} P_{ii}^{(n-k)} \\
&= 1 + \sum_{k=1}^{\infty} e^{tk} f_{ii}^{(k)} \sum_{n=k}^{\infty} e^{t(n-k)} P_{ii}^{(n-k)} \\
&= 1 + P_i(t) \sum_{k=1}^{\infty} e^{tk} f_{ii}^{(k)} = 1 + P_i(t) F_i(t).
\end{aligned}$$

所以 $P_i(t) = \dfrac{1}{1-F_i(t)}$。$t<0$ 及正常返性保证了级数的绝对收敛及和号交换的合法。于是有

$$(1-e^t) P_i(t) = \frac{1-e^t}{1-F_i(t)}. \tag{3.14}$$

分别对(3.14)式左右两端求极限：

$$\lim_{t\uparrow 0^-} (1-e^t) P_i(t) = \lim_{t\uparrow 0^-} \frac{\sum_{n=0}^{\infty} e^{tn} P_{ii}^{(n)}}{\sum_{n=0}^{\infty} e^{tn}}$$

$$= \lim_{t \uparrow 0^-} \lim_{k \to \infty} \frac{\sum_{n=0}^{k} e^{tn} P_{ii}^{(n)}}{\sum_{n=0}^{k} e^{tn}} = \lim_{k \to \infty} \frac{\sum_{n=0}^{k} P_{ii}^{(n)}}{k} = \lim_{k \to \infty} P_{ii}^{(k)}.$$

上面 $\lim_{t \uparrow 0^-}$ 的含义是 t 从 0 的左边趋于 0。最后一个等式利用了微积分中的 Stolz 定理。它告诉我们非负收敛序列的极限和他们算术平均的极限是一样的。而在(3.14)的右边，由 L'Hopital 法则

$$\lim_{t \uparrow 0^-} \frac{1-e^t}{1-F(t)} = \lim_{t \uparrow 0^-} \frac{-e^t}{-\sum_{n=0}^{\infty} n e^{tn} f_{ii}^{(n)}} = \frac{1}{\sum_{n=0}^{\infty} n f_{ii}^{(n)}} = \frac{1}{\mu_i},$$

定理得证。

一个正常返非周期的状态也称作是遍历的(ergodic)。定理 3.3 c)告诉我们对遍历状态 i 有 $\lim_{n \to \infty} P_{ii}^{(n)} = \frac{1}{\mu_i}$，其中 μ 是常返时的期望。

至于不同状态间的 n 步转移概率 $P_{ji}^{(n)}$ 有

推论 3.2 如果状态 i 是遍历的则对所有 $j \to i$

有 $$\lim_{n \to \infty} P_{ji}^{(n)} = \lim_{n \to \infty} P_{ii}^{(n)} = \frac{1}{\mu_i}$$

证明留给读者作为练习。

直接求 P^n 的每一元素 $P_{ij}^{(n)}$ 的极限并不是一件容易的事。而求条件期望 ET_i 也并不方便。它涉及求一系列条件概率 $f_{ii}^{(n)}$。因而人们寻求更简捷的方法来处理极限分布。

定义 3.6 Markov 链有转移概率阵 $P = \|P_{ij}\|$。一个概率分布 $\{\pi_i, i \geq 0\}$ 如果满足 $\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}$ 则称作是这一 Markov 链的平稳分布。

很容易看出如果过程的初始状态 X_0 有概率分布 $\pi = \{\pi_j, j \geq 0\}$ 也即 $P\{X_0 = j\} = \pi_j$。则有

$$P\{X_1 = j\} = \sum_{i=0}^{\infty} P\{X_1 = j | X_0 = i\} P\{X_0 = i\} = \sum_{i=0}^{\infty} \pi_i P_{ij} = \pi_j.$$

并由归纳法可得

$$P\{X_n=j\}=\sum_{i=0}^{\infty}P\{X_n=j|X_{n-1}=i\}P\{X_{n-1}=i\}=\sum_{i=0}^{\infty}\pi_i P_{ij}=\pi_j.$$

于是对所有 n，X_n 有相同的分布，再加上 Markov 性立即可知对任何 $k\geqslant 0$，$X_n, X_{n+1}, \cdots, X_{n+k}$ 的联合分布不依赖于 n。这就是说，$\{X_n, n\geqslant 0\}$ 作为随机过程是平稳的。平稳分布 π 由此得名。有了这一概念我们可以讨论重要的

定理 3.4 若一个不可约 Markov 链中的所有状态都是遍历的，则对所有 i 极限 $\lim_{n\to\infty}P_{ij}^{(n)}=\pi_j$ 存在且 $\pi=\{\pi_j, j\geqslant 0\}$ 为平稳分布。也即

$$\sum_j \pi_j=1, \ \pi_j>0, \tag{3.15}$$

$$\sum_i \pi_i P_{ij}=\pi_j. \tag{3.16}$$

反之，若一个不可约 Markov 链存在一个平稳分布，即满足(3.15)及(3.16)式。则这个 Markov 链的所有状态都是遍历的。而且该平稳分布就是这一 Markov 链的极限分布。即对任何 i 有

$$\lim_{n\to\infty}P_{ij}^{(n)}=\pi_j \tag{3.17}$$

证明 由定理 3.3 c) 及推论 3.2 已知对遍历状态 j

$$\lim_{n\to\infty}P_{ij}^{(n)}=\frac{1}{\mu_j}$$

且 $\mu_j<\infty$。于是 $\frac{1}{\mu_i}>0$，记之为 π_j。由 $\sum_{j=0}^{\infty}P_{ij}^{(n)}=1$ 通过积分号取极限可以证明 $\sum_{j=0}^{\infty}\pi_j\leqslant 1$。再利用 Chapman-Kolmogrov 方程

$$P_{ij}^{(n)}=\sum_{k=0}^{\infty}P_{ik}^{(n-1)}P_{kj},$$

并令 $n\to\infty$，一方面可推出对所有 M

$$\pi_j=\lim_{n\to\infty}P_{ij}^{(n)}=\lim_{n\to\infty}\sum_{k=0}^{\infty}P_{ik}^{(n-1)}P_{kj}\geqslant\lim_{n\to\infty}\sum_{k=0}^{M}P_{ik}^{(n-1)}P_{kj}$$

$$=\sum_{k=0}^{M}\pi_k P_{kj}.$$

也即 $\pi_j \geqslant \sum_{k=0}^{\infty} \pi_k P_{kj}$。另一方面如若 π_j 严格大于 $\sum_{k=0}^{\infty} \pi_k P_{kj}$ 则有

$$\sum_{j=0}^{\infty} \pi_j > \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi_k P_{kj} = \sum_{k=0}^{\infty} \pi_k \sum_{j=0}^{\infty} P_{kj} = \sum_{k=0}^{\infty} \pi_k.$$

前面已证的 $\sum_{j=0}^{\infty} \pi_j \leqslant 1$ 保证了交换求和运算的合理性及 $\sum_j \pi_j$ 的收敛。于是导出了矛盾。由反证法知必有(3.16)式成立。即

$$\pi_j = \sum_{k=0}^{\infty} \pi_k P_{kj}.$$

不断地将(3.16)式代入上面的 π_k 即可得到

$$\pi_j = \sum_{k=0}^{\infty} \pi_k P_{kj}^{(n)},$$

再令 $n \to \infty$ 即有 $\pi_j = \sum_{k=0}^{\infty} \pi_k \pi_j = \pi_j \sum_{k=0}^{\infty} \pi_k$。由 $\pi_j > 0$，推出 $\sum_{k=0}^{\infty} \pi_k = 1$。$\pi = \{\pi_j, j \geqslant 0\}$ 确为平稳分布。

反过来，假定一个不可约的 Markov 链存在一个平稳分布 $\{\pi_j\}$，则由归纳法可知对任何 n，

$$\pi_j = \sum_{k=0}^{\infty} \pi_k P_{kj}^{(n)}. \tag{3.18}$$

由链的不可约性，链中各状态属同一类型。而由 $\pi_j > 0$ 及 $\sum_j \pi_j = 1$ 知它们不可能为瞬过的或零常返的，因为这时将有 $\lim_{n \to \infty} P_{kj}^{(n)} = 0$。因而只能有 $\lim_{n \to \infty} P_{kj}^{(n)} = \frac{1}{\mu_j}$。在(3.18)中令 $n \to \infty$，即有

$$\pi_j = \sum_{k=0}^{\infty} \pi_k \frac{1}{\mu_j} = \frac{1}{\mu_j}$$

从而证明了这个极限分布正是平稳分布。

例 3.8 续 在例 3.8 中(3.15)及(3.16)式化为

$$\begin{cases} \pi_0 + \pi_1 = 1, \\ \pi_0 P_{00} + \pi_1 P_{10} = \pi_0, \\ \pi_0 P_{01} + \pi_1 P_{11} = \pi_1. \end{cases}$$

代入 P_{ij} 的值有

$$\begin{cases}\pi_0+\pi_1=1,\\(1-\alpha)\pi_0+\beta\pi_1=\pi_0,\\\alpha\pi_0+(1-\beta)\pi_1=\pi_1,\end{cases}$$

其解为 $\pi_0=\dfrac{\beta}{\alpha+\beta}$，$\pi_1=\dfrac{\alpha}{\alpha+\beta}$。这是我们在例 3.8 的计算中已知的。殊途而同归。

在实际应用中 $\{\pi_j\}$ 有两种解释：一是作为 $P_{ij}^{(n)}$ 的极限分布它告诉我们在过程的长期运行中不论初始状态 i 是什么，经过一段时期后发现过程处于状态 j 的概率就是 π_j。另一解释是 π_j 也代表了就长期而言过程访问 j 的次数在总时间中的平均份额或比例。这可从下面的推理中看出。设

$$I_n=\begin{cases}1,\ 若\ X_n=j,\\0,\ 其他,\end{cases}$$

那么 m 步转移中访问 j 的次数所占比例为 $\dfrac{1}{m}\sum\limits_{n=0}^{m-1}I_n$。若初始状态为 $X_0=i$。则这一比例的条件期望为

$$E\left\{\frac{1}{m}\sum_{n=0}^{m-1}I_n\,|\,X_0=i\right\}=\frac{1}{m}\sum_{n=0}^{m-1}P\{X_n=j\,|\,X_0=i\}$$

$$=\frac{1}{m}\sum_{n=0}^{m-1}P_{ij}^{(n)},$$

当 $m\to\infty$ 时，仍由 Stolz 定理

$$\lim_{m\to\infty}\frac{1}{m}\sum_{n=0}^{m-1}P_{ij}^{(n)}=\lim_{m\to\infty}P_{ij}^{(m)}=\pi_j,$$

这就支持了第二种解释。

有关周期的 Markov 链以及 Markov链是可约的情况有关极限分布的计算就更为复杂。限于篇幅不在此作详细的讨论。

3.4 分支过程

分支过程（Branching process）的思想早在 1874 年就由

F. Galton 首先提出。当时他研究的是家族姓氏的消失问题。这种数学模型是一类特殊的 Markov 链。它在生物遗传、原子核的链锁反应中都有应用。模型的大意是在第 0 代(常称为祖先)有一个个体，他所繁衍的第一代子女数为一随机变量.可能是 0，1，2，…而每个第一代个体又再衍生子孙。整个群体可能兴旺，也可能消亡。比如在原子连锁反应中原子由中子的轰击而分裂，分裂出的中子又有机会撞击别的原子而发生连锁反应。因而上一代撞击所产生中子的多少会直接影响下一代所能产生中子的数量。记 X_0 为第 0 代祖先的数目、一般假定为 1，并记 X_n 为第 n 代后裔的大小。Markov 链 $\{X_n, n\geqslant 0\}$ 就称作分支过程。记 Z_i 为第 n 代中第 i 个个体所繁衍的后代数，则有 $X_{n+1}=\sum_{i=1}^{X_n} Z_i$。也就是说 X_{n+1} 为 X_n 个随机变量 Z_i 的随机和。我们可以简化地假定各代个体的繁衍能力是相同的，而且同一代个体的繁衍是相互独立的。这时 Z_i 独立同分布，$P(Z_1=k)=p_k$，且记 $EZ_1=\mu_1$，$\mathrm{Var}Z_1=\sigma^2$。这时就可以算出 Markov 链 $\{X_n, n\geqslant 0\}$ 的转移概率为

$$P_{ij}=P\{X_{n+1}=j\,|\,X_n=i\}=P\left\{\sum_{k=1}^{i} Z_k=j\right\}.$$

利用第一章的(1.20)式我们可求出 X_{n+1} 的均值与方差满足.

$$EX_{n+1}=EX_n\cdot EZ_1. \tag{3.19}$$

$$\mathrm{Var}\,X_{n+1}=EX_n\,\mathrm{Var}\,Z_1+\mathrm{Var}\,X_n(EZ_1)^2. \tag{3.20}$$

由(3.19)式不断地迭代即有

$$EX_{n+1}=EX_n\cdot\mu=EX_{n-1}\cdot\mu^2=EX_1\cdot\mu^n=\mu^{n+1}$$

这里要注意到第 0 代只有一个个体所以 X_1 与 Z_1 同分布。类似地通过(3.20)式可知

$$\begin{aligned}
\mathrm{Var}X_{n+1}&=\mu^n\cdot\sigma^2+\mathrm{Var}\,X_n\cdot\mu^2\\
&=\mu^n\cdot\sigma^2+\mu^2[\mu^{n-1}\cdot\sigma^2+\mathrm{Var}(X_{n-1})\mu^2]\\
&=\mu^n\cdot\sigma^2+\mu^{n+1}\cdot\sigma^2+\mu^4[\mu^{n-2}\sigma^2+\mathrm{Var}(X_{n-2})\cdot\mu^2]\\
&=\cdots=\sigma^2\mu^n[1+\mu+\cdots+\mu^n]
\end{aligned}$$

$$=\begin{cases}\sigma^2\mu^n\dfrac{(1-\mu^{n+1})}{1-\mu}, & \mu\neq 1,\\(n+1)\sigma^2, & \mu=1.\end{cases}$$

由(3.19)，(3.20)可以看出μ的大小对于整个群体和家族的繁衍存亡至关重要。当$\mu<1$，且n很大时EX_n与$\text{Var } X_n$都趋于0。由著名的Chebyshev不等式马上可以推出对任意给的$\varepsilon>0$有

$$P\{|X_n-EX_n|>\varepsilon\}\leqslant\frac{\text{Var } X_n}{\varepsilon^2},$$

也即当$n\to\infty$时可以证明$P\{|X_n-0|>\varepsilon\}\to 0$或者说$X_n$会以概率趋于0。群体终将消亡。而当$\mu\geqslant 1$群体消亡的概率就是一个较为复杂的问题。但这是分支过程中的令人感兴趣的量。从Markov链状态分类的观点看群体消亡就是在某个时刻n过程X_n进入吸收状态，而群体消亡的概率就是过程最终进入吸收状态的概率。至于在中子轰击原子裂变的物理问题中$p_0=P(Z_1=0)=0$，即粒子不会消失。而p_1则为原子没被击中或不分裂的概率。这时人们最感兴趣的就不是消亡概率而是粒子数增加的速率了。

我们用生成函数来研究消亡概率。记过程X_n的生成函数为$\phi_n(s)$，X_1的生成函数则直接记为$\phi(s)$。即$\phi(s)=\sum_{k=0}^{\infty}p_ks^k$，其中$p_k=P(X_1=k)$，$k=0,1,2,\cdots$。由第一章(1.24)式随机和$X_{n+1}$的生成函数可以表为$\phi_{n+1}(s)=\phi_n(\phi(s))$或$\phi_{n+1}(s)=\phi(\phi_n(s))$。记$\pi_n$为由单个祖先开始家族将在第$n$代之前消亡的概率。则$\pi_n=P\{X_n=0\}=\phi_n(0)$。我们所感兴趣的整个群体终将消亡的概率为当$n\to\infty$时$\pi_n$的极限。当$p_0=1$，家族不能发端。而当$p_0=0$它将永不会消亡。所以一般假定$0<p_0<1$。这时对$s\in(0,1)$区间

$$\phi'(s)=\frac{\mathrm{d}}{\mathrm{d}s}\left(\sum_{j=0}^{\infty}p_js^j\right)=\sum_{j=1}^{\infty}p_j\cdot j\cdot s^{j-1}>0.$$

$\phi(s)$是(0.1)区间上的单调函数，而$\phi(0)=p_0>0$，所以

$$\pi_{n+1}=\phi_{n+1}(0)=\phi_n[\phi(0)]>\phi_n(0)=\pi_n.$$

因此π_n随n的单调上升，极限$\lim\limits_{n\to\infty}\pi_n$存在。记之为$\pi$就是群体消亡的概率。由$\phi_{n+1}(s)=\phi(\phi_n(s))$知$\pi_{n+1}=\phi(\pi_n)$，再令$n\to\infty$即有$\pi=\phi(\pi)$。于是通过求解函数方程$\phi(s)=s$就可以求出$\pi$。事实上还可以严格证明下面的

定理 3.5 对分支过程X_n，若$p_0>0$，$p_0+p_1<1$则有

a) 群体消亡概率π是方程$\phi(s)=s$的最小正解。其中$\phi(s)=\sum\limits_{j=0}^{\infty}p_js^j$，$\{p_j\}$是$X_1$与$Z_1$的概率分布。

b) $\pi=1$，当且仅当$\mu\leqslant1$，其中$\mu=EZ_1$。

证明 我们仅证明a)，b)的证明要涉及一点凸函数的知识。有兴趣的读者不妨作为一道数学分析的证明题来练习。问题是若$\phi(s)$为$(0,1)$开区间上的严格凸函数。则1是函数方程$\phi(s)=s$的最小正解当且仅当$\phi'(1)\leqslant1$。为证a)，设ϖ为方程$\phi(s)=s$的任一正根，则有$\pi_1=\phi(0)<\phi(\varpi)=\varpi$由$\phi_{n+1}(s)=\phi(\phi_n(s))$及归纳法可得$\pi_{n+1}=\phi(\pi_n)<\phi(\varpi)=\varpi$所以$\pi=\lim\limits_{n\to\infty}\pi_n\leqslant\varpi$。

定理的b)告诉我们当$\mu\leqslant1$时群体必然消亡。而当$\mu>1$时群体有正概率$1-\pi$生长到无穷。

例 3.9 假定某生物群体中每个个体所衍生的下一代的个数服从参数为λ的Poisson分布。于是有

$$p_k=P(Z=k)=\frac{\lambda^k}{k!}e^{-\lambda},\ k=0,1,2,\cdots.$$

因为i个独立的参数为λ的Poisson随机变量的和仍为Poisson随机变量，而参数为λi。所以分支过程X_n的转移概率

$$p_{ij}=P\{X_{n+1}=j\mid X_n=i\}=\frac{(\lambda i)^j}{j!}e^{-\lambda i}.$$

而Poisson变量的生成函数为$\phi(s)=\exp\{\lambda(s-1)\}$，曲线$e^{\lambda(s-1)}$与直线在$(0,1)$区间仅有一个交点。这个交点$\pi$就是群体消亡的概率，而当$\lambda>1$时群体有正概率$1-\pi$增长到无穷大。而当$\lambda\leqslant1$时$\pi=1$群体必然消亡。比如取$\lambda=2$，$\pi\doteq0.2$。

分支过程还可以有各种推广．比如一个已成年的个体产生 Z 个下一代，Z 有生成函数 $\phi(s)$，而未成年的个体则以概率 p 长成年，以概率 q，$q=1-p$，仍停留在未成年状态，所以无法繁衍后代。于是产生后代的数目这一随机变量的生成函数为 $(1-p)+p\phi(s)$。这时每一个体繁衍后代的期望数为 $p\phi'(1)$，方差为 $p[\phi''(1)+\phi'(1)-(\phi'(1))^2]$。如果群体中有 k 个个体而且相互独立地繁衍后代则产生的后代数会有生成函数 $(q+p\phi(s))^k$。这是一类多重分支过程的例子。人们还可以考虑更为复杂的带移民的，或带杀伤的群体生长模型。它们可以在生物，人口及社会科学中有广泛的应用。限于篇幅不能一一加以介绍。

3.5 连续时间 Markov 链

3.5.1 连续时间 Markov 链

顾名思义，连续时间 Markov 链的要点仍是 Markov 性．所谓 Markov 性就是给定了过程在目前的状态其未来的发展与过去的状况是独立的。它与离散时间 Markov 链不同的只是把时间指标参数从离散的 0，1，2，… 改为连续的实数 t。一般有 $t\geqslant 0$。而过程的取值或所处的状态仍以整数来标志。许多重要的概率模型比如 Poisson 过程，生灭过程等都是连续时间 Markov 链。

定义 3.6 若对所有 s，$t\geqslant 0$ 和任何非负整数 i，j，$x(u)$，$0\leqslant u\leqslant s$。随机过程 $X(t)$，$t\geqslant 0$ 满足

$$P\{X(t+s)=j|X(s)=i,X(u)=x(u),0\leqslant u<s\}$$
$$=P\{X(t+s)=j|X(s)=i\}, \tag{3.21}$$

则称作是连续时间 Markov 链。

式子 (3.21) 所表示的就是 Markov 性。如果更进一步 $P\{X(t+s)=j|X(s)=i\}$ 是与 s 无关的则这一 $X(t)$ 又称为是有

平稳转移概率的连续时间 Markov 链。它将是我们的主要研究对象。这时平稳转移概率又可简单地记为 $P_{ij}(t)$。过程的初始分布仍记为 $p_i = P\{X(0)=i\}$。于是我们仍然有

命题 3.5 连续时间 Markov 链的转移概率 $P_{ij}(t)$ 和 p_i 完全确定了过程的所有联合分布。

证明 由 Markov 性对任何 $0<t_1<t_2<\cdots<t_n$ 和状态 i_n，$n=0,1,2,\cdots$，

$$P\{X(t_n)=i_n, X(t_{n-1})=i_{n-1}, \cdots, X(0)=i_0\}$$
$$=P\{X(t_n)=i_n|X(t_{n-1})=i_{n-1}, \cdots, X(0)=i_0\}\cdot$$
$$P\{X(t_{n-1})=i_{n-1}|X(t_{n-2})=i_{n-2}, \cdots, X(0)=i_0\}\cdots P\{X_0=i_0\}$$
$$=P_{i_{n-1},i_n}(t_n-t_{n-1})\cdots P_{i_0,i_1}(t_1)p_{i_0}.$$

现在的问题是一组对所有 $t\geqslant 0$ 有定义的函数 $P_{ij}(t)$ 应满足什么条件才能充当一个 Markov 链的转移概率？有些条件是显而易见的，比如，

$$P_{ij}(t)\geqslant 0, \sum_j P_{ij}(t)=1. \tag{3.22}$$

此外，若当过程从状态 i 转移到状态 j 用去了时间 $t+\tau$ 则在时刻 τ 过程必然会处于某个状态 k。于是

$$P\{X(t+\tau)=j, X(\tau)=k|X(0)=i\}$$
$$=P\{X(t+\tau)=j|X(\tau)=k, X(0)=i\}$$
$$P\{X(\tau)=k|X(0)=i\}$$
$$=P_{kj}(t)P_{ik}(\tau).$$

所以 $P_{ij}(t)$ 应满足连续形式的 Chapman-Kolmogrov 方程

$$P_{ij}(t+\tau)=\sum_k P\{X(t+\tau)=j, X(\tau)=k|X(0)=i\}$$
$$=\sum_k P_{ik}(\tau)P_{kj}(t). \tag{3.23}$$

另外再添加一个使过程在有限的时间内不至于转移无穷多次的条件，就是说过程不能刚到某状态就瞬即离去。这要求对任何状态 i

$$\lim_{\tau\downarrow 0} P\{X(t+\tau)=i|X(t)=i\}=\lim_{\tau\downarrow 0} P_{ii}(\tau)=1. \tag{3.24}$$

事实上，可以证明

定理 3.6 函数 $P_{ij}(t)$ 能够作为无瞬即转移的 Markov 过程的转移概率函数的充分必要条件是它满足条件(3.22)，(3.23)和(3.24)。

连续时间 Markov 链与离散时间 Markov 链不同的是当我们考虑过程时不仅关心它会转移到什么状态去，转移的机会有多大。我们还关心过程在某个状态逗留的时间有多长。如果我们假定在时刻 0 过程处于状态 i 并在 s 个单位时间内都留在 i 没有离去，那么在以后 t 单位时间内过程仍不离开 i 的概率有多大？由 Markov 性知过程在 $[s, s+t]$ 时间段留在 i 的概率（实际上是条件概率）是与过程至少在状态 i 逗留 t 个单位时间是一样的。过程好像忘记了时刻 s 之前的过去。于是若记 τ_i 为过程在转移之前逗留于状态 i 的时间，则应有 $P\{\tau_i > s+t | \tau_i > s\} = P\{\tau_i > t\}$。若此式对所有 $s, t > 0$ 成立则称随机变量 τ_i 有无记忆性。由概率论的知识不难推出 τ_i 服从指数分布。记它的参数为 ν_i。这样一分析我们对连续时间 Markov 链的了解就清楚了。连续时间 Markov 链是有以下两条性质的随机过程：a）过程从某一状态 i 转移到其他状态之前在 i 逗留的时间服从以 ν_i 为参数的指数分布；b）过程离开状态 i 在下一次转移时转入状态 j 的概率为 P_{ij}，这里 $\sum\limits_{j \neq i} P_{ij} = 1$。请注意 a）中的参数 ν_i 依赖 i 这说明逗留时间分布在不同状态可能不一样，但都同属于指数分布类型。有了这两条性质就能比较容易地了解连续时间 Markov 链。一般，不考虑 $\nu_i = \infty$ 的情形。当 $\nu_i = 0$ 时状态 i 便成了吸收态，一旦进入便永远无法转移。因为 ν_i 代表了过程离开状态 i 的速率而 P_{ij} 是过程发生转移时从 i 转入 j 的概率。所以，若 $j \neq i$ 记 $q_{ij} = \nu_i P_{ij}$ 则 q_{ij} 代表了过程从 i 到 j 的转移率。

例 3.10 考虑 Poisson 过程 $\{X(t), t \geq 0\}$。在时间区间 $[t, t+\tau]$ 上到达的粒子数服从以 $\lambda\tau$ 为参数的 Poisson 分布。由独立增量性，对 $0 \leq t_1 < \cdots < t_n < t$，

$$
\begin{aligned}
P\{X(t+\tau)=k+j \mid X(t)=j, X(t_n)=j_n, \cdots, X(t_1)=j_1\} \\
&= \frac{(\lambda\tau)^k}{k!}e^{-\lambda\tau}=P\{X(t+\tau)-X(t)=k\} \\
&= P\{X(t+\tau)=k+j \mid X(t)=j\}.
\end{aligned}
$$

所以 $X(t)$ 为有平稳转移概率的连续时间 Markov 链。其转移概率为

$$
P_{ij}(t)=\begin{cases} 0, & j<i, \\ \dfrac{(\lambda\tau)^{j-i}}{(j-i)!}e^{-\lambda\tau}, & j\geqslant i. \end{cases}
$$

对 Poisson 过程当 $j\neq i+1$ 时 $q_{ij}=0$，而当 $j=i+1$ 时 $P_{i,i+1}=1$，$\nu_i=\lambda$ 所以转移率 $q_{i,i+1}=\lambda$。两次转移之间的间隔时间服从参数为 λ 的指数分布。这是在第二章中早已熟知的。Poisson 过程是理解连续时间 Markov 链的基本范例。

在分析研究连续时间 Markov 链时常用的办法是无穷小分析。利用 Markov 性对很小的 h 求出有关 $P_{ij}(h)$ 的关系式，或转移概率应当满足的微分方程组。然后依据边界条件对它求解。在第二章分析 Poisson 过程时我们已用过类似的方法。

3.5.2 纯生过程

纯生过程是 Poisson 过程的自然推广。是考虑一个生物群体时在保证良好环境、充足食物、无死亡、无外迁的理想条件下的生长模型。而新生一个个体的概率是和群体当前的大小（即过程所处状态）有关的并与时间长短成比例.在流行病学的研究中纯生过程可以用来建立新病例增殖的模型。考虑一串正数 $\{\lambda_k\}$ $k=0, 1, 2, \cdots$，记 $X(t)$ 为 $(0, t]$ 时间段中新生的个体数。当 $X(t)$ 满足以下四条假定时就称作是一个纯生过程。

$$
\left.\begin{array}{ll}
\text{(i)} & X(0)=0, \\
\text{(ii)} & P\{X(t+h)-X(t)=1 \mid X(t)=k\}=\lambda_k h+o(h), \\
\text{(iii)} & P\{X(t+h)-X(t)=0 \mid X(t)=k\}=1-\lambda_k h+o(h), \\
\text{(iv)} & P\{X(t+h)-X(t)<0 \mid X(t)=k\}=0.
\end{array}\right\} \quad (3.25)
$$

(3.25)式中(ii)，(iii)两条假定若采用转移概率的记号就是

$$P_{k,k+1}(h)=\lambda_k h+o(h),\ P_{k,k}(h)=1-\lambda_k h+o(h).$$

简记 $P_{0n}(t)=P\{X(t)=n\mid X(0)=0\}$ 为 $P_n(t)$。于是由全概率公式及假定(iv)知 $P_n(t)$ 应满足关系式

$$\begin{aligned}P_n(t+h)&=\sum_{k=0}^{\infty}P_k(t)P(X(t+h)=n\mid X(t)=k)\\&=\sum_{k=0}^{\infty}P_k(t)P\{X(t+h)-X(t)=n-k\mid X(t)=k\}.\end{aligned}$$

再利用(ii)及(iii)

$$\begin{aligned}P_n(t+h)=P_n(t)&[1-\lambda_n h+o(h)]\\&+P_{n-1}(t)[\lambda_{n-1}h+o(h)]+o(h)\end{aligned}$$

再移项整理即有微分方程组

$$\left.\begin{aligned}&P_0'(t)=-\lambda_0 P_0(t)\\&P_n'(t)=-\lambda_n P_n(t)+\lambda_{n-1}P_{n-1}(t),\ n\geqslant 1\end{aligned}\right\}\qquad(3.26)$$

其中第一个方程的解是一眼就可以看出的即 $P_0(t)=e^{-\lambda_0 t}$，$t>0$。后一方程则稍费点事，需引入 $Q_n(t)=e^{\lambda_n t}P_n(t)$，利用(3.26)可知 $Q_n'(t)=e^{\lambda_n t}\lambda_{n-1}P_{n-1}(t)$ 而 $Q_n(0)=P_n(0)=0$。对等式两边从 0 到 t 积分即有

$$Q_n(t)=\int_0^t e^{\lambda_n x}\lambda_{n-1}P_{n-1}(x)\mathrm{d}x,$$

也即

$$P_n(t)=\lambda_{n-1}e^{-\lambda_n t}\int_0^t e^{\lambda_n x}P_{n-1}(x)\mathrm{d}x,\ n=1,2,\cdots.$$

如果 $\lambda_1\neq\lambda_0$ 则有

$$P_1(t)=\lambda_0\left(\frac{1}{\lambda_1-\lambda_0}e^{-\lambda_0 t}+\frac{1}{\lambda_0-\lambda_1}e^{-\lambda_1 t}\right),$$

而当 $\lambda_1=\lambda_0$ 时

$$P_1(t)=\lambda_0 te^{-\lambda_0 t}.$$

逐次迭代对一般的 n 可得到 $P_n(t)$ 的表达形式。在 λ_i 两两不同时为

$$P_n(t)=\prod_{i=0}^{n-1}\lambda_i\left[\sum_{i=0}^{n}c_{i,n}e^{-\lambda_i t}\right],$$

其中 $c_{i,n}=\dfrac{1}{\prod\limits_{k\neq i}(\lambda_k-\lambda_i)}$。应该说方程组(3.26)的求解已不是本课程的任务了。不熟悉的读者可参看微分方程的书籍。

记纯生过程在某一状态 i 的逗留时间为 T_i 并用 $W_k=\sum\limits_{i=0}^{k-1}T_i$ 表示第 k 个新生命诞生的时刻。则有

$$P_n(t)=P\{W_n\leqslant t<W_{n+1}\}.$$

可以证明 $T_i, i=0,1,\cdots$ 是相互独立的参数为 λ_i 的指数分布。例如首次新生的等待时间 T_0 的分布 $P(T_0\leqslant t)=1-P_0(t)=1-e^{-\lambda_0 t}$ 确是参数为 λ_0 的指数分布。这些参数 λ_k 表明过程在状态 i 逗留时间的长短。而 $\dfrac{1}{\lambda_k}$ 表示在状态 i 的平均逗留时间。可以证明为了保证 $\sum\limits_{n=0}^{\infty}P_n(t)=1$，这时 $X(t)=\infty$ 的概率为 $1-\sum\limits_{n=0}^{\infty}P_n(t)$ 等于 0，有一个充分条件是 $\sum\limits_{n=0}^{\infty}\dfrac{1}{\lambda_n}=\infty$。也即当群体增长到无穷大的平均时间 $\left(\sum\limits_{n=0}^{\infty}\dfrac{1}{\lambda_n}\right)$ 为无穷大时，过程在有限时刻 t 为 ∞ 的概率为 0。因而过程必然在时刻 t 时有个体数为某个 n，这又是 $\sum\limits_{n=0}^{\infty}P_n(t)=1$ 的概率含义。因而条件 $\sum\limits_{n=0}^{\infty}\dfrac{1}{\lambda_n}=\infty$ 也是必要的。

例 3.11 Yule 过程。这是一类特殊的纯生过程，在物理和生物学中有广泛的应用。假定在长度为 h 的小时间区段中群体的每一个成员都有概率 $\beta h+o(h)$ 产生一个新的个体并假定成员的行为是相互独立的。于是有

$$\begin{aligned}&P\{X(t+h)-X(t)=1\mid X(t)=n\}\\&=\binom{n}{1}[\beta h+o(h)][1-\beta h+o(h)]^{n-1}\\&=n\beta h+o(h).\end{aligned}$$

也即 $P_{n,n+1}(h)=n\beta h+o(h)$，或 $\lambda_n=n\beta$。此时方程(3.26)在 $X(0)=1$ 的情形下变为类似的

$$P_n'(t) = -\beta[nP_n(t) - (n-1)P_{n-1}(t)],\ n = 1, 2, \cdots.$$

其中 $P_n(t) = P\{X(t) = n | X(0) = 1\}$。初始条件为 $P_1(0) = 1$, $P_n(0) = 0$, $n = 2, 3, \cdots$ 方程的解为

$$P_n(t) = e^{-\beta t}(1 - e^{-\beta t})^{n-1}, \quad n \geq 1. \qquad (3.27)$$

这是一个以 $p = e^{-\beta t}$ 为参数的几何分布。在关于一般纯生过程解的表达式中将 $\lambda_k = k\beta$ 代入即有

$$c_{i,n} = (-1)^{i-1}/\beta^{n-1}(i-1)!\,(n-i)!.$$

所以

$$\begin{aligned} P_n(t) &= \beta^{n-1}(n-1)! \sum_{i=1}^{n} c_{i,n} e^{-i\beta t} \\ &= \sum_{i=1}^{n} \frac{(n-1)!}{(i-1)!\,(n-i)!} (-1)^{i-1} e^{-i\beta t} \\ &= e^{-\beta t} \sum_{j=0}^{n-1} \binom{n-1}{j} (-e^{\beta t})^j \\ &= e^{-\beta t}(1 - e^{-\beta t})^{n-1} \end{aligned}$$

正好与(3.27)式吻合。

除纯生过程外，人们也可以考虑纯灭过程，这时生物群体死亡率或设备元件的损坏率可以用 μ_k 表示。在与(3.25)类似的假定下建立有关的微分方程，并求出相应的解。与将齐次的 Poisson 过程推广到非齐次的 Poisson 过程相类似人们还可以考虑 λ 是时间的函数的时变 Yule 过程。在纯生过程中还可进一步假定出现一个新生个体的概率为 $\lambda(k, t)h + o(h)$。当取 $\lambda(k, t) = \dfrac{\lambda + \lambda a k}{1 + \lambda a t}$ 时就是著名的 Polya 过程。$\lambda(k, t)$ 是 t 的线性函数的倒数。这些都无法一一详细讨论了。

3.6 生 灭 过 程

3.6.1 生灭过程(birth and death process)

纯生过程和纯灭过程的自然推广是可以同时考虑 $X(t)$ 增加

减少的生灭过程。模型的假定是这样的：$X(t)$ 是状态 0，1，2，…上的 Markov 链其转移概率 $P_{ij}(t)$ 是平稳的，即对所有 $u\geqslant 0$ 有 $P\{X(t+u)=j\mid X(u)=i\}=P_{ij}(t)$。此外还假定

$$\left.\begin{array}{l}\text{(i) } P_{i,i+1}(h)=\lambda_i h+o(h),\ i\geqslant 0\\ \text{(ii) } P_{i,i-1}(h)=\mu_i h+o(h),\ i\geqslant 1,\\ \text{(iii) } P_{ii}(h)=1-(\lambda_i+\mu_i)h+o(h),\\ \text{(iv) } P_{ii}(0)=1,P_{ij}(0)=0,\\ \text{(v) } \mu_0=0,\mu_i>0\quad i\geqslant 1;\lambda_i>0\quad i\geqslant 0,\end{array}\right\}\tag{3.28}$$

其中 λ_i 和 μ_i 分别称作新生率和死亡率。过程的机制是当它处于状态 i 时只能转到 $i+1$ 或 $i-1$ 这两个状态。前者称为新生、后者代表死亡。而过程在状态 i 的逗留时间服从 $\nu_i=\lambda_i+\mu_i$ 的指数分布。转移率 $q_{i,i+1}=\lambda_i$，$q_{i,i-1}=\mu_i$ 其他均为 0。于是从关系式 $q_{ij}=\nu_i P_{ij}\quad i\neq j$ 立即可知 $P_{i,i+1}=\dfrac{\lambda_i}{\lambda_i+\mu_i}=1-P_{i,i-1}$。这说明当给定时刻 t 过程发生一次转移时转移是新生的概率为 $\dfrac{\lambda_i}{\lambda_i+\mu_i}=\dfrac{\lambda_i}{\nu_i}$ 恰是新生率与总转移率 ν_i 的比值。

在建立随机模型时还要注意的一点是并非任一组 $\{\lambda_i\}$，$\{\mu_i\}$ 都可以充当生灭过程的新生率与死亡率。为使有唯一的 Markov 过程其转移概率满足 $P_{ij}(t)\geqslant 0$，$\sum\limits_{j=0}^{\infty}P_{ij}(t)\leqslant 1$ 及 Chapman-Kolmogorov 方程，这些参数应当满足条件

$$\sum_{n=0}^{\infty}\frac{1}{\lambda_n\theta_n}\sum_{k=0}^{n}\theta_k=\infty,\tag{3.29}$$

其中 $\theta_0=1$，$\theta_n=\prod\limits_{i=1}^{n}\dfrac{\lambda_{i-1}}{\mu_i}$，$n=1,2,\cdots$。好在对通常的实际问题这一条件不难满足。

3.6.2 Kolmogorov 向后向前微分方程

从 Chapman-Kolmogorov 方程出发我们可以得到两个有关

转移概率的微分方程组。分别称为 Kolmogorov 向后或向前微分方程。

定理 3.6 对生灭过程 $X(t)$ 的转移概率 $P_{ij}(t)$ 有 Kolmogorov 向后微分方程

$$\begin{cases} P'_{ij}(t) = \mu_i P_{i-1,j}(t) - (\lambda_i + \mu_i) P_{ij}(t) \\ \qquad\qquad + \lambda_i P_{i+1,j}(t), \ i \geqslant 0 \\ P'_{0j}(t) = -\lambda_0 P_{0j}(t) + \lambda_0 P_{1j}(t). \end{cases} \tag{3.30}$$

和 Kolmogrov 向前微分方程

$$\begin{cases} P'_{ij}(t) = \mu_{j+1} P_{i,j+1}(t) - (\lambda_j + \mu_j) P_{ij}(t) + \lambda_{j-1} P_{i,j-1}(t) \\ P'_{i0}(t) = -\lambda_0 P_{i0}(t) + \mu_1 P_{i1}(t). \end{cases} \tag{3.31}$$

证明 由 Chapman-Kolmogorov 方程知

$$\begin{aligned} P_{ij}(t+h) &= \sum_{k=0}^{\infty} P_{ik}(h) P_{kj}(t) \\ &= P_{i,i-1}(h) P_{i-1,j}(t) + P_{ii}(h) P_{ij}(t) + P_{i,i+1}(h) P_{i+1,j}(t) \\ &\quad + \sum_k{}^* P_{ik}(h) P_{kj}(t). \end{aligned}$$

和式 $\sum_k^*$ 是对所有 $k \neq i-1, i, i+1$ 求和。由假定(i)、(ii)、(iii)可以估计出

$$\begin{aligned} \sum_k{}^* P_{ik}(h) P_{kj}(t) &\leqslant \sum_k{}^* P_{ik}(h) \\ &= 1 - [P_{i,i-1}(h) + P_{ii}(h) + P_{i,i+1}(h)] \\ &= 1 - [\mu_i h + o(h) + 1 - (\lambda_i + \mu_i) h + o(h) + \lambda_i h + o(h)] \\ &= o(h). \end{aligned}$$

所以，

$$\begin{aligned} P_{ij}(t+h) &= \mu_i h \ P_{i-1,j}(t) + [1 - (\lambda_i + \mu_i) h] P_{ij}(t) \\ &\quad + \lambda_i h P_{i+1,j}(t) + o(h) \end{aligned}$$

将 $P_{ij}(t)$ 移项除以 h 并令 $h \downarrow 0$ 即有

$$P'_{ij}(t) = \mu_i P_{i-1,j}(t) - (\lambda_i + \mu_i) P_{ij}(t) + \lambda_i P_{i+1,j}(t).$$

(3.30)式得证。如果在 Chapman-Kolmogorov 方程中将 h 和 t 的位置互换则有

$$P_{ij}(t+h)=\sum_{k=0}^{\infty}P_{ik}(t)P_{kj}(h).$$

于是

$$\frac{P_{ij}(t+h)-P_{ij}(t)}{h}=\sum_{k\neq j}P_{ik}(t)\frac{P_{kj}(h)}{h}$$

$$-P_{ij}(t)\left(\frac{1-P_{jj}(h)}{h}\right).$$

在极限运算与求和可以交换的条件满足时不难利用假定(i)、(ii)、(iii)算出

$$P'_{ij}(t)=\sum_{k\neq j}P_{ik}(t)\cdot\lim_{h\to0}\frac{P_{kj}(h)}{h}-P_{ij}(t)\lim_{h\to0}\frac{1-P_{jj}(h)}{h}$$

$$=\lambda_{j-1}P_{i,j-1}(t)-(\lambda_j+\mu_j)P_{ij}(t)+\mu_{j+1}P_{i,j+1}(t).$$

向后方程是把时间区间$(0, t+h)$分为$(0, h)$和$[h, t+h)$来考查状态的转移。$P_{kj}(t)$是给定$X(h)=k$时$X(t+h)=j$的条件概率。h很小所以在$(0, t+h)$中远在过去。而向前方程则把时间区间$(0, t+h)$分为$(0, t)$和$[t, t+h)$。给定的条件是$X(t)=k$，所考察的是未来长度为h的小区间。向后向前由此得名。

例 3.12 带移民a的线性增长和线性死亡模型。取$\lambda_n=\lambda n+a$，$\mu_n=\mu n$，其中λ，μ，$a>0$这在人口问题中是常见的。我们最感兴趣的是在时刻t的期望人口数$E\{X(t)\}$。利用向前微分方程将λ_i，μ_i代入即有

$$P'_{i0}(t)=-aP_{i0}(t)+\mu P_{i1}(t)$$

$$P'_{ij}(t)=[\lambda(j-1)+a]P_{i,j-1}(t)-[(\lambda+\mu)j+a]P_{ij}(t)$$
$$+\mu(j+1)P_{i,j+1}(t),\quad j\geqslant1$$

记$E[X(t)]=m(t)=\sum_{j=1}^{\infty}jP_{ij}(t)$，于是$m(t)$应满足方程

$$m'(t)=a+(\lambda-\mu)m(t).$$

设初始条件为$X(0)=i$，则$m(0)=i$。当$\lambda=\mu$时容易求出$m(t)=at+i$，当$\lambda\neq\mu$时

$$m(t)=\frac{a}{\lambda-\mu}\{e^{(\lambda-\mu)t}-1\}+ie^{(\lambda-\mu)t},$$

当$\lambda \geqslant \mu$时，若$t\to\infty$，$m(t)\to\infty$而当$\lambda<\mu$时$m(t)$的极限为$\frac{a}{\mu-\lambda}$。经过长时期后人口的平均数将趋于统计平衡。

至于连续时间 Markov 链的极限分布也可以有与离散情形相类似的讨论.并可以求出平衡分布.有兴趣的读者参看文献[2]。而当平衡分布存在时其所对应随机变量的期望就是上例中的极限。

到目前为至所讨论的还仅限于$X(t)$只能取有限个或可列无穷多个状态的情形。再推而广之就可以讨论连续时间的 Markov 过程。其中最著名的例子是本书第五章要讨论的 Brown 运动。在结束本章之前，我们给出

定义 3.7 随机过程$X(t)$，$t\geqslant 0$，若对任何s，$t\geqslant 0$，$0\leqslant u\leqslant s$，其条件概率分布函数满足

$$P\{X(t+s)\leqslant a \mid X(s)=x, X(u)=x(u), 0\leqslant u<s\} \\ =P\{X(t+s)\leqslant a \mid X(s)=x\}. \tag{3.32}$$

则称作是一个 Markov 过程。

(3.32)所代表的仍然是 Markov 性质。

习 题 三

1. 对 Markov 链X_n，$n\geqslant 0$，试证条件

$$P\{X_{n+1}=j \mid X_0=i_0, \cdots, X_{n-1}=i_{n-1}, X_n=i\} \\ =P\{X_{n+1}=j \mid X_n=i\},$$

等价于对所有时刻n，m及所有状态i_0，$\cdots$，i_n，j_1，$\cdots$，j_m

有
$$P\{X_{n+1}=j_1, \cdots, X_{n+m}=j_m \mid X_0=i_0, \cdots, X_n=i_n\} \\ =P\{X_{n+1}=j_1, \cdots, X_{n+m}=j_m \mid X_n=i_n\}.$$

2. 考虑状态 0，1，2 上的一个 Markov 链X_n，$n\geqslant 0$ 它有转移概率矩阵

$$
P=\begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array}\begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left\|\begin{array}{ccc} 0.1 & 0.2 & 0.7 \\ 0.9 & 0.1 & 0 \\ 0.1 & 0.8 & 0.1 \end{array}\right\| \end{array},
$$

初始分布为 $p_0=0.3$，$p_1=0.4$，$p_2=0.3$。试求概率

$$P\{X_0=0, X_1=1, X_2=2\}.$$

3. 信号传送问题。信号只有 0，1 两种。分为多个阶段传输。在每一步上出错的概率为 α。$X_0=0$ 是送出的信号，而 X_n 是在第 n 步接收到的信号。假定 X_n 为一 Markov 链它有转移概率矩阵 $P_{00}=P_{11}=1-\alpha$，$P_{01}=P_{10}=\alpha$，$0<\alpha<1$。试求

(a) 两步均不出错的概率：

$$P\{X_0=0, X_1=0, X_2=0\},$$

(b) 试求两步传送后收到正确信号的概率，

(c) 试求五步之后传送无误的概率：

$$P\{X_5=0|X_0=0\}.$$

4. A，B 两罐总共装着 N 个球。作如下试验：在时刻 n 先从 N 个球中等概率地任取一球。然后从 A，B 两罐中任选一个，选中 A 的概率为 p，选 B 的概率为 q。之后再将选出的球放入选好的罐中。设 X_n 为每次试验时 A 罐中的球数。试求此 Markov 过程的转移概率矩阵。

5. 重复掷币一直到连续出现两次正面为止。假定钱币是均匀的，试引入以连续出现次数为状态空间的 Markov 链。并求出平均需要掷多少次试验才可以结束。

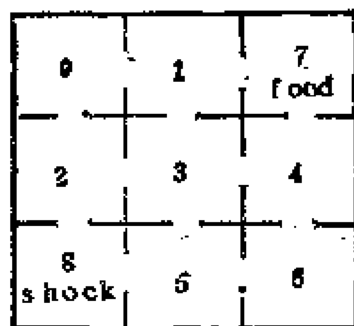


图 3.1

6. 迷宫问题。将小家鼠放入迷宫中作动物的学习试验，如图 3.1 所示。在迷宫的第 7 号小格内放有美味食品而第 8 号小格内则是电击捕鼠装置。假定当家鼠位于某格时有 k 个出口可以离去，则它总是随机地选择一个，概率为 $1/k$。并假定每一次家鼠只能跑到相邻的小格

去。令过程 X_n 为家鼠在时刻 n 时所在小格的号码。试写出这一 Markov 过程的转移概率阵，并求出家鼠在遭到电击前能找到食物的概率。

7. 记 Z_i，$i=1, 2, \cdots$ 为一串独立同分布的离散随机变量。$P\{Z_1=k\}=p_k\geqslant 0, k=0, 1, 2, \cdots, \sum_{k=0}^{\infty} p_k=1$。记 $X_n=Z_n$，$n=0, 1, 2, \cdots$。试求过程 X_n 的转移概率阵。

8. 对习题 7 中的 Z_i。令 $X_n=\max\{Z_1, \cdots, Z_n\}$ $n=1, 2, \cdots$ 并约定 $X_0=0$。X_n 是否为 Markov 链？如果是，其转移概率阵是什么？

9. $f_{ij}^{(n)}$ 代表从 i 出发在 n 步转移时首次到达 j 的概率，试证明

$$P_{ij}^{(n)}=\sum_{k=0}^{n} f_{ij}^{(k)} P_{jj}^{(n-k)}.$$

10. 对习题 7 中的 Z_i。若定义 $X_n=\sum_{i=1}^{n} Z_i$，$n=1, 2, \cdots X_0=0$，试证 X_n 为 Markov 链，并求其转移概率矩阵。

11. 一 Markov 链有状态 0，1，2，3 和转移概率矩阵

$$\begin{Vmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{Vmatrix}.$$

试求 $f_{00}^{(n)}$，$n=1, 2, 3, 4, 5, \cdots$，其中 $f_{ii}^{(n)}$，由

$$P\{X_n=i, X_k\neq i, k=1, \cdots, n-1 \mid X_0=i\} \text{ 定义。}$$

12. 在成败型的重复试验中。每次试验结果为成功(S)或失败(F)。同一结果相继出现称为一个游程 (run)，比如一串结果 FSSFFFSF 中共有两个成功游程，三个失败游程。设成功概率为 p，失败概率为 $q=1-p$。记 X_n 为 n 次试验后成功游程的长度。

（若第 n 次试验失败则 $X_n=0$）。试证 $\{X_n, n=1, 2, \cdots\}$ 为一 Markov 链。并确定其转移概率阵。记 T 为返回状态 0 的时间。试求 T 的分布及均值，并由此对这一Markov链的状态进行分类。

13. 试证向各方向游动的概率相等的对称随机游动在二维时是常返的，而在三维时却是瞬过的。

14. 某厂商对该厂生产的同类产品的三种型号调查顾客的消费习惯。并把它们归结为 Markov 链模型。记顾客消费习惯在 A，B，C 三种型号间的转移概率阵分别为下列四种。请依这些转移阵所提供的信息对厂家提出关于 A，B 两种型号的咨询意见。

（Ⅰ）$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$，　（Ⅱ）$\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$，

（Ⅲ）$\begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$，　（Ⅳ）$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$。

15. 考虑一有限状态的 Markov 链。试证明

(a) 至少有一个状态是常返的，

(b) 任何常返状态必定是正常返的。

16. 考虑一生长与灾害模型。这类 Markov 链有状态 0，1，2，…。当过程处于状态 i 时它既可能以概率 p_i 转移到 $i+1$（生长）也能以概率 $q_i=1-p_i$ 落回状态 0（灾害）。而从状态"0"又必然"无"中生有。即 $P_{01}=1$。

(a) 试证所有状态为常返的条件是

$$\lim_{n\to\infty}(p_1p_2p_3\cdots p_n)=0,$$

(b) 若此链为常返的，试求其为零常返的条件.

17. 试计算转移概率阵

$$P=\begin{array}{c}\\0\\1\\2\end{array}\begin{array}{c}\begin{array}{ccc}0&1&2\end{array}\\\left\|\begin{array}{ccc}\frac{1}{2}&\frac{1}{2}&0\\\frac{1}{3}&\frac{1}{3}&\frac{1}{3}\\\frac{1}{6}&\frac{1}{2}&\frac{1}{3}\end{array}\right\|\end{array}$$

的极限分布。

18. 假定在逐日的天气变化模型中，每天的阴晴与前两天的状况关系很大。于是可考虑4状态的Markov链：接连两晴天，一晴一阴，一阴一晴，以及接连两阴天分别记为(S,S)，(S,C)，(C,S)，和(C,C)。该链的转移概率阵为

	(S,S)	(S,C)	(C,S)	(C,C)
(S,S)	0.8	0.2	0	0
(S,C)	0	0	0.4	0.6
(C,S)	0.6	0.4	0	0
(C,C)	0	0	0.1	0.9

试求这一Markov链的平稳分布。并求出长期平均的晴朗天数。

19. 某人有M把伞并在办公室和家之间往返。如某天他在家时(办公室时)下雨了而且家中(办公室)有伞他就带一把伞去上班(回家)。不下雨时他从不带伞。如果每天与以往独立地早上(晚上)下雨的概率为p。试定义一$M+1$状态的Markov链以研究他被雨淋湿的机会。

20. 血液培养在时刻0从一个红细胞开始一分钟之后红细胞死亡可能出现下面几种情况：以概率1/4再生2个红细胞，以

$\frac{1}{2}$的概率再生1个红细胞和一个白细胞，也有$\frac{1}{4}$的概率产生2个白细胞。再过一分钟后每个红细胞以同样的规律再生下一代而白细胞则不再生，并假定每个细胞的行为是独立的。

(a) 从培养开始 $n+\frac{1}{2}$ 分钟不出现白细胞的概率是多少？

(b) 整个培养过程停止的概率是多少？

21．分支过程中一个体产生后代的分布为 $p_0=q$，$p_1=p(p+q=1)$ 试求第 n 代总体的均值和方差及群体消亡的概率。如产生后代的分布为 $p_0=\frac{1}{4}$，$p_1=\frac{1}{2}$，$p_2=\frac{1}{4}$ 及 $p_0=\frac{1}{8}$，$p_1=\frac{1}{2}$，$p_2=\frac{1}{4}$，$p_3=\frac{1}{8}$ 试回答同样的问题。

22．若单一个体产生后代的分布为 $p_0=q$，$p_2=p$，$p+q=1$。并假定过程开始时的祖先数为1试求分支过程第3代总数的分布。

23．一连续时间 Markov 链有0和1两个状态，在状态0和1的逗留时间服从参数为 $\lambda>0$ 及 $\mu>0$ 的指数分布。试求在时刻0从状态0起始，t 时刻后过程处于状态0的概率 $P_{00}(t)$。

24．在习题23中如果 $\lambda=\mu$。定义 $N(t)$ 为过程在 $[0,t]$ 中改变状态的次数，试求 $N(t)$ 的概率分布。

25．记 $X(t)$ 为纯生过程，且有

$P\{X(t+h)-X(t)=1|X(t)$ 为奇数$\}=\alpha h+o(h)$，

$P\{X(t+h)-X(t)=1|X(t)$ 为偶数$\}=\beta h+o(h)$。

及 $X(0)=0$。试分别求事件“$X(t)$ 为偶数”及“$X(t)$ 为奇数”的概率。

26．考虑状态 $0,1,\cdots,N$ 上的纯生过程 $X(t)$，假定 $X(0)=0$ 以及 $\lambda_k=(N-k)\lambda$，$k=0,1,\cdots,N$。其中 λ_k 满足

$$P\{X(t+h)-X(t)=1|X(t)=k\}=\lambda_k h+o(h)。$$

试求 $P_n(t)=P(X(t)=n)$。这是新生率受群体总数反馈作用的

例子。

27. 在某化学反应中，由分子 A 与 B 发生反应而产生分子 C。假定在时间小 h 之内一个分子 A 与 B 接近到能发生化学反应的概率是与 h 及 A，B 当前的分子数成正比。假定在反应开始时 A，B 分子数相同，并记过程 $X(t)$ 为 A 分子在时刻 t 的数目。试建立其随机过程模型。

28. 有无穷多个服务员的排队系统：假定顾客以参数为 λ 的 Poisson 过程到达。而服务员的数量巨大，可理想化为无穷多个。顾客一到就与别的顾客相独立地接受服务，并在时间 h 内完成服务的概率近似为 αh。记 $X(t)$ 为在时刻 t 正接受服务的顾客总数。试建立此过程的转移机制的模型。

29. 一个由 N 个部件组成的循环装置。从 C_1，C_2，…到 C_N 顺时针排列。第 k 个部件会持续工作一段时间，其分布为以 λ_k 为参数的指数分布，一旦它停止工作，顺时针方向的下一个元件就立即接替它开始运行。假定各部件及同一部件的不同次运行都是相互独立的。记 $X(t)$ 为时刻 t 正在运行的部件的序号。试写出模型及转移概率所满足的微分方程。当 $N=2$，$\lambda_1=\lambda_2=1$，初始状态为 1 时试求解 $P_{11}(t)$ 及 $P_{12}(t)$。

30. 试写出纯生过程的 Kolmogorov 向前微分方程。在初始条件 $P_{ii}(0)=1$ 下试求出 $P_{ii}(t)$ 及 $P_{ij}(t)$ 应满足的方程。特别对 $\lambda_j=j\lambda$ 的 Yule 过程求出 $P_{ij}(t)$ 的明显表达式。

31. 两个通讯卫星放入轨道。每一个卫星的工作寿命为以 μ 为参数的指数分布。一旦失效就再放射一颗新卫星替换它。所需的准备及发射时间为一以 λ 为参数的指数分布。记 $X(t)$ 为时刻 t 时在轨道中工作的卫星数，假定这是一个状态空间为 $\{0，1，2\}$ 的连续时间 Markov 链模型。试建立 Kolmogorov 向前及向后微分方程。

第4章 平稳过程

平稳过程 $X=\{X(t),\ t\in T\}$ 是其概率性质在时间平移下不变的随机过程。这一思想抓住了没有固定时间(空间)起点的物理系统中的最自然的现象，因而在通讯理论、天文学、生物学、生态学和经济学等各个领域中获得了广泛的应用。

由平稳性可以导出许多理论中的重要结论，本章主要讨论遍历性定理，谱表示理论和预报理论。

4.1 定义和例子

以下设 T 为具有如下性质的下标集合：若 $t_1,\ t_2\in T$，则 $t_1+t_2\in T$。通常 T 取如下几种集合之一：

(i) $T=\{0,1,2,\cdots\}$,

(ii) $T=\{0,\pm1,\pm2,\cdots\}$,

(iii) $T=\{t:\ t\geqslant 0\}$,

(iv) $T=\{t:-\infty<t<\infty\}$.

直观上，下标集合可以理解为时间(当然，T也可以表示空间位置或其他)。

定义 4.1 设 $X=\{X(t),\ t\in T\}$ 为一随机过程，若对任意正整数 k 及 T 中任意 k 个时刻 $t_1<t_2<\cdots<t_k$,及 T 中的 h，有

$$\{X(t_1),\cdots,X(t_k)\}\overset{d}{=}\{X(t_1+h),\cdots,\ X(t_k+h)\},\qquad(4.1)$$

则随机过程 X 称为严平稳过程。这儿“d”表示等式两边 k 维随机向量的分布相同。

注：如果 $T=\{0,\ \pm1,\cdots\}$，我们一般把 X 称为随机序列。如果X还是严平稳的，则称为严平稳序列，以下不再一一说明。

设 $X=\{X(t),\ t\in T\}$ 为严平稳过程，由定义如果均值函数 $m(t)=EX(t)$ 存在，则必为常数，即 $m(t)=m,\ t\in T$。同样，如果方差函数存在，则 $\mathrm{Var}(X(t))=E(X(t)-m)^2$ 也是一个常数。设 $s,\ t\in T$，（不妨设 $s<t$），由平稳性，其协方差函数

$$E(X(t)-m)(X(s)-m)=E(X(t-s)-m)(X(0)-m).$$

等式右端只依赖于时间差 $t-s$。若记

$$R(h)=E(X(h)-m)(X(0)-m),$$

则

$$E(X(t+h)-m)(X(t)-m)=R(h),$$

即协方差函数仅与时间差有关，而与起点无关。当然，由定义知 $\mathrm{Var}(X(t))=R(0)$。此外，我们称

$$\rho(v)=R(v)/\sigma^2=R(v)/R(0)$$

为平稳过程 X 的相关函数或自相关函数。由概率论中相关系数性质易知 $\rho(0)=1$ 及 $|\rho(v)|\leqslant 1$。

下面是平稳过程的几个简单例子。

例 4.1 通讯理论中的电脉冲常用一个平稳过程来描述。当然，在任一物理系统中，产生一个信号总需要一个短暂时间。由于这段时间与信号长度相比很小，所以可以用平稳过程来描述。此外，在电子通讯理论中，电位和电流常用一个复随机变量表示，因此我们还可能遇到复值的平稳过程。

例 4.2 星星、星云、植物和动物的空间（或平面）分布经常用平稳过程来描述，这儿 T 可以是欧氏空间，球的表面或平面等等。

例 4.3 海洋波的高度也可能作为平稳过程来处理，其中 T 是球面经度和纬度坐标。

对于一个严平稳过程而言，要求它的所有有限维分布都与起点无关，这无疑是一个相当严格的条件，而且在实际中也很难验证。而从实际物理系统来看，过程的前二阶矩已能反映该过程的许多重要的概率性质。因此引入下面较弱的平稳过程的概念。

定义 4.2 设 $X=\{X(t),\ t\in T\}$ 为一实值随机过程，如果

对 $\forall t\in T$，$EX^2(t)<\infty$，$EX(t)=m$ 以及协方差函数 $E(X(t)-m)(X(s)-m)$ 仅与 $t-s$ 有关，则称 X 为宽平稳随机过程。

一般来说，这两个过程是互不包含的，即严平稳由于不一定有二阶矩而不必是宽平稳；反之，宽平稳由于其有限维联合分布可能不满足 (4.1) 式而不一定为严平稳。但只要过程的二阶矩存在，则严平稳过程一定是宽平稳过程。

例 4.4 设 Z 为非零随机变量，考虑如下两个随机过程：$X=\{X(t)=tZ,t\in T\}$，$Y=\{Y(t)=Z,t\in T\}$。试考察这两个随机过程的平稳性。

解 对随机过程 Y 而言，由于在时刻 t 的取值与 t 无关，故必然是严平稳的。如果 $EZ^2<\infty$，则 Y 也是宽平稳的。至于随机过程 X，由于其一维分布与 t 有关，所以必然不是严平稳的。其次，即使 $EZ^2<\infty$，如果 $EZ\neq 0$，则 $EX(t)=tEZ$，与 t 有关，因此不是宽平稳。如果 $EZ=0$，由于 $EX(t)X(s)=tsEZ^2$ 与 t,s 有关，故仍不是宽平稳。

一个自然的问题是是否存在一类过程，使这两个平稳过程的定义等价呢？答案是肯定的，为此我们引入如下的定义。

定义 4.3 设 $G=\{G(t),-\infty<t<\infty\}$ 为一随机过程，如果对任一整正数 k 以及任意 k 个时刻 $t_1\leqslant t_2\leqslant\cdots\leqslant t_k$，$(G(t_1),G(t_2),\cdots,G(t_k))$ 的联合分布为 k 维正态分布，则称 G 为高斯 (Gauss) 过程。

我们知道，k 维正态分布完全由协方差矩阵和均值向量所唯一确定，而这些量仅与它们的前二阶矩有关，所以对 Gauss 过程而言，两个平稳的定义是等价的。

我们以下主要研究宽平稳过程。为方便起见，我们就称宽平稳过程为平稳过程。

例 4.5 （平稳白噪声序列），设 $X_n,n=0,1,\cdots$ 为一列两两不相关的随机变量序列，满足 $EX_n=0$，$EX_n^2=\sigma^2$，$n=0,1,\cdots$，且 $EX_mX_n=0$ 当 $m\neq n$。则 $X=\{X_n,n=0,1,\cdots\}$ 为平稳序列。

这是因为协方差函数

$$EX_mX_n=\begin{cases}\sigma^2, & m=n,\\ 0, & m\neq n,\end{cases}\qquad(4.2)$$

仅与 $m-n$ 有关。

例 4.6 （三角多项式过程）一些关于平稳过程的有意义的例子可以通过考虑某种有随机振幅的三角函数表达式得到。设 A 和 B 为同分布的随机变量，均值为 0，方差为 σ^2，且 A 和 B 不相关，即 $EAB=0$。对 $[0,\pi]$ 中一个固定的频率 ω，定义

$$X_t=A\cos\omega t+B\sin\omega t,$$

则 $X=\{X(t):-\infty<t<\infty\}$ 为平稳过程。这是由于

$$EX_t=\cos\omega tEA+\sin\omega tEB=0.$$

由 $EAB=0$ 得

$$\begin{aligned}EX_{t+\tau}X_t&=E(A\cos(t+\tau)\omega+B\sin(t+\tau)\omega)(A\cos t\omega+B\sin t\omega)\\&=EA^2\cos(t+\tau)\omega\cos t\omega+EB^2\sin(t+\tau)\omega\sin t\omega\\&=\sigma^2\cos\tau\omega.\end{aligned}\qquad(4.3)$$

即协方差函数仅与时间差 τ 有关，故为平稳过程。

更一般地，我们可以作如下的推广。设 $A_0, A_1, \cdots, A_m$；$B_0,B_1,\cdots,B_m$ 为两两不相关的随机变量，且 $EA_i=EB_j=0$，$EA_i^2=EB_i^2=\sigma_i^2, 0\leqslant i\leqslant m, 0\leqslant j\leqslant m$。记 $\omega_0,\omega_1,\cdots,\omega_m$ 为 $[0,\pi]$ 中不同的频率。定义

$$X_t=\sum_{k=0}^{m}(A_k\cos t\omega_k+B_k\sin t\omega_k),$$

则显然有 $EX_t=0$。利用诸 A_i,B_j 不相关，用得出 (4.3) 式类似的讨论可知

$$EX_{t+\tau}X_t=\sum_{k=0}^{m}\sigma_k^2\cos\tau\omega_k\qquad(4.4)$$

仅与 τ 有关，故 $\{X_t,-\infty<t<\infty\}$ 为平稳过程。注意到 $A_k\cos t\omega_k$ 和 $B_k\sin t\omega_k$ 表示圆频率为 ω_k 而随机振幅为 A_k 和 B_k 的简谐振动在时刻 t 时质点的位置。由此看出，当振动的随机振幅两两不

相关时，经迭加而生成的过程仍为平稳过程。上面（4.4）式中，若令 $\sigma^2=\sigma_0^2+\cdots+\sigma_m^2$，$p_k=\sigma_k^2/\sigma^2$，则可改写为

$$R(\tau)=\sigma^2\sum_{k=0}^{m}p_k\cos\tau\omega_k,$$

这儿 p_k 表示频率 ω_k 在协方差函数中的贡献份额。注意到 $\{p_k, 0\leqslant k\leqslant m\}$ 是离散概率密度函数。若以 F 记在 ω_k 有跳为 p_k 的分布函数，则上式可记为

$$R(\tau)=\sigma^2\int_0^{\pi}\cos\tau\omega\,\mathrm{d}F(\omega). \tag{4.5}$$

（4.5）式启发我们可以把离散频率推广到连续频率状态：

$$R(\tau)=\sigma^2\int_0^{\pi}\cos\tau\omega\,\mathrm{d}F(\omega),$$

其中 F 为定义在 $[0,\pi]$ 上某个连续频率分布函数，后面将会看到这种推广确是可行的。特别当 F 为 $[0,\pi]$ 上的均匀分布而 τ 取整值 $0,1,\cdots$ 时

$$R(\tau)=\sigma^2\int_0^{\pi}\frac{1}{\pi}\cos\tau\omega\,\mathrm{d}\omega=\begin{cases}\sigma^2, & \tau=0,\\ 0, & \tau\neq 0,\end{cases}$$

这就是例 5 中的协方差函数。

例 4.7 （滑动平均序列）设 $\{\varepsilon_n, n=0,\pm1,\pm2,\cdots\}$ 为一列不相关的有相同均值 m 和方差 σ^2 的随机变量。设 $a_1,\cdots,a_k$ 为任意 k 个实数。考虑由下式定义的序列：

$$X_n=a_1\varepsilon_n+a_2\varepsilon_{n-1}+\cdots+a_k\varepsilon_{n-k+1},\quad n=0,\ \pm1,\cdots$$

我们有

$$EX_n=m(a_1+\cdots+a_k)$$

记 $\xi_j=\varepsilon_j-m$，则由 ε_j 的两两不相关知协方差函数

$$\begin{aligned}
&r(n+\tau,n)\\
&=E[(X_n-m(a_1+\cdots+a_k))(X_{n+\tau}-m(a_1+\cdots+a_k))]\\
&=E(a_1\xi_n+a_2\xi_{n-1}+\cdots+a_k\xi_{n-k+1})\\
&\quad\cdot(a_1\xi_{n+\tau}+a_2\xi_{n+\tau-1}+\cdots+a_k\xi_{n+\tau-k+1})\\
&=\begin{cases}\sigma^2(a_k a_{k-\tau}+\cdots+a_{\tau+1}a_1), & \text{若 } 0\leqslant\tau\leqslant k-1,\\ 0, & \text{若 } \tau\geqslant k,\end{cases}
\end{aligned}$$

即协方差函数仅与时间间隔 τ 有关，故 $\{X_n, n=0,\pm1,\cdots\}$ 为平稳过程。如果取 $a_j=1/\sqrt{k}$，$j=1,\cdots,k$，则

$$R(\tau)=\begin{cases}\sigma^2(1-\tau/k), & \tau\leqslant k-1,\\ 0 & \tau\geqslant k.\end{cases}$$

例 4.8 （随机电报信号）设信号流 $X=\{X(t),\ t\geqslant0\}$ 为一随机过程，且对每个 t 有

$$P(X(t)=I)=P(X(t)=-I)=\frac{1}{2},$$

而在$[t,\ t+\tau]$时间内正负号变化的次数 N 服从速率为 λ 的 Poisson过程，即

$$P(N(t)=k)=\mathrm{e}^{-\lambda t}(\lambda t)^k/k!,\qquad \lambda>0.$$

试讨论信号流 X 的平稳性。

解 显然，$EX(t)=0$，$t\geqslant0$。为计算协方差函数，设 $\tau\geqslant0$。注意到 $X(t+\tau)$与 $X(t)$的乘积只能取 I^2 和 $-I^2$ 两个值。故

$$EX(t+\tau)X(t)=I^2\cdot P(\text{信号在}[t,\ t+\tau]\text{内变号偶数次})$$
$$-I^2\cdot P(\text{信号在}[t,t+\tau]\text{内变号奇数次}).$$

但

$$P(\text{信号在}[t,t+\tau]\text{内变号偶数次})$$
$$=\sum_{k=0}^{\infty}P(N(t+\tau)-N(t)=2k)$$
$$=\sum_{k=0}^{\infty}\mathrm{e}^{-\lambda\tau}(\lambda\tau)^{2k}/(2k)!=\mathrm{e}^{-\lambda\tau}\mathrm{ch}(\lambda\tau),$$

$$P(\text{信号在}[t,t+\tau]\text{内变号奇数次})$$
$$=\sum_{k=0}^{\infty}P(N(t+\tau)-N(t)=2k+1)$$
$$=\sum_{k=0}^{\infty}\mathrm{e}^{-\lambda\tau}(\lambda\tau)^{2k+1}/(2k+1)!=\mathrm{e}^{-\lambda\tau}\mathrm{sh}(\lambda\tau).$$

所以

$$EX(t+\tau)X(t)=I^2\sum_{k=0}^{\infty}\mathrm{e}^{-\lambda\tau}(-\lambda\tau)^k/k!=I^2\mathrm{e}^{-\lambda\tau}\mathrm{e}^{-\lambda\tau}=I^2\mathrm{e}^{-2\lambda\tau},$$

当 $\tau<0$ 时，考虑$[t+\tau,t]$内信号变化次数同样可得

$$EX(t+\tau)X(t)=I^2 e^{-2\lambda|\tau|}$$

即协方差函数只与时间间隔 τ 有关，因此随机电报信号流是平稳过程。

在平稳过程中，一般常见的过程是周期平稳过程。

定义 4.4 设 $X=\{X(t), t\in T\}$为平稳过程，如果存在正常数 κ 使

$$X(t+\kappa)=X(t),$$

则 X 称为周期平稳过程，κ 为过程的周期。如果 X 是周期平稳过程，则其协方差函数也是周期函数，且与过程有相同的周期。因此

$$\begin{aligned}R(\tau+\kappa)&=E(X(t+\tau+\kappa)-m)(X(t+\kappa)-m)\\&=E(X(t+\tau)-m)(X(t)-m)=R(\tau).\end{aligned}$$

还可以考虑复平稳过程。其定义与定义 4.2 类似，差别一是把 $X(t)$ 改为复值，二是把协方差函数定义为 $E(X(t)-m)\cdot\overline{(X(s)-m)}$，这儿$\overline{X}$表示$X$的共轭。

例 4.9 （周期振动） 设 $X(t)=X\cdot f(t)$，其中 X 为实随机变量，满足 $EX=0$，$EX^2=\sigma^2$，$f(t)$为一个非随机的复值函数。考虑复值过程 $X=\{X(t), -\infty<t<\infty\}$。试证明 X 为平稳过程的充分必要条件是 $f(t)=Ce^{j(\lambda t+\theta)}$，其中 $j=\sqrt{-1}$，C，λ，θ 为常数。

证 由 $EX=0$ 知 $EX(t)=0, -\infty<t<\infty$。如果 $f(t)=Ce^{j(\lambda t+\theta)}$，则

$$\begin{aligned}EX(t+\tau)\overline{X(t)}&=EX^2C^2e^{j(\lambda(t+\tau)+\theta)}e^{-j(\lambda t+\theta)}\\&=\sigma^2C^2e^{j\lambda\tau}\end{aligned}$$

故 X 为平稳过程。反之，设 $X(t)$是一个平稳过程，则

$$R(\tau)=EX(t+\tau)\overline{X(t)}=EX^2f(t+\tau)\bar{f}(t)=\sigma^2f(t+\tau)\bar{f}(t),$$

即 $f(t+\tau)\bar{f}(t)$ 与 t 无关。取$\tau=0$，我们有

$$|f(t)|^2=C^2=R(0),$$

故 $f(t)=Ce^{j\psi(t)}$，这儿 $\psi(t)$ 为一实函数。由此

$$f(t+\tau)\bar{f}(t)=C^2 e^{j(\psi(t+\tau)-\psi(t))}$$

与 t 无关，因此

$$\frac{d(\psi(t+\tau)-\psi(t))}{dt}=0,$$

即 $\frac{d\psi(t+\tau)}{dt}=\frac{d\psi(t)}{dt}$ 对一切 t 均成立。因而 $\psi'(t)$ 为一常数，记为 λ。则 $\psi(t)=\lambda t+\theta$，因而有

$$f(t)=Ce^{j(\lambda t+\theta)}.$$

4.2 遍历性定理

考虑如下两个平稳过程：

$Z=\{X_n, n=0,1,2,\cdots\}$，其中 $\{X_n\}$ 为独立同分布随机变量序列，$EX_n^2<\infty$；$EX_n=m$，$n=0,1,\cdots$。

$Y=\{Y_n=Y, n=0,1,2,\cdots\}$，其中 Y 是随机变量，$EY^2<\infty$。

这两个随机过程是平稳过程中的二个极端情况，可以用这两个过程来阐述不同平稳过程之间的差异。例如，由大数法则知对 X 过程来说，我们有

$$\frac{1}{n}(X_0+X_1+\cdots+X_{n-1})$$

以概率收敛于常数 m。但对 Y 过程来说

$$\frac{1}{n}(Y_0+Y_1+\cdots+Y_{n-1})=Y,$$

即经过对时间的平均后，随机性没有任何改变。于是自然产生这样的问题：对平稳过程加上什么条件后，对时间的平均值可以等于过程的均值？这一问题称为平稳过程的遍历性（Ergodic）问题。这是平稳过程研究中的一个重要课题。其重要性可以从如下粗略的分析中看出。对平稳过程 X，重要的是确定它的均值 m 以及协方差函数 $R(\tau)$，由于 $EX(t)=m$，为估计 m，我们必须对随机过程 X 作大量观察，以 $X_i(t)$ 记第 i 次观察中时刻 t 的值。

$i=1, 2, \cdots, n$。则大数法则告诉我们可以用 $\hat{m}_n=\frac{1}{n}(X_1(t)+\cdots+X_n(t))$ 来估计 m。同样，为了估计协方差函数 $R(\tau)$，也可以用

$$\hat{R}_n(\tau)=\frac{1}{n}\sum_{k=1}^{n}(X_k(t+\tau)-\hat{m}_n)(X_k(t)-\hat{m}_n)$$

来估计。但对随机过程作多次观察一般来说很难做到。比较容易的是作了一次观察，获得一条样本路径，我们希望由这一次观察来估计 m 和 $R(\tau)$。不言而喻，对于一般的随机过程这是不太可能的。但对于平稳过程，只要加上一些很一般的条件，比如观察的时间足够长，就可以从一次观察中获得 m 和 $R(\tau)$ 的较好的估计。这就是有名的遍历性定理。为导出这一定理，我们需要如下的定义。

定义 4.5 设 $X=\{X(t), -\infty<t<\infty\}$ 为一平稳过程（或序列），若*

$$\overline{X}=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}X(t)\,\mathrm{d}t=m \tag{4.6}$$

或

$$\overline{X}=\lim_{N\to\infty}\frac{1}{2N+1}\sum_{k=-N}^{N}X(k)=m, \tag{4.7}$$

则称 X 的均值有遍历性。如果

$$\overline{R}(\tau)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}(X(t)-m)(X(t+\tau)-m)\mathrm{d}t=R(\tau) \tag{4.8}$$

或

$$\overline{R}(\tau)=\lim_{N\to\infty}\frac{1}{2N+1}\sum_{k=-N}^{N}(X(k)-m)(X(k+\tau)-m)=R(\tau), \tag{4.9}$$

则称 X 的协方差函数有遍历性。若随机过程（或序列）的均值和协方差函数都有遍历性，则称此随机过程有遍历性。

遍历性又称各态历经性。直观上可以这样理解。考虑只有有

* 这儿的极限均指均方极限

限个状态的平稳序列$\{X_n, n=0,1,\cdots\}$，其状态$E=\{e_1,\cdots,e_b\}$。则$m=EX_n$是各个状态的加权平均。若令$A_N=\{X_n: -N\leqslant n\leqslant N\}$，则遍历性告诉我们，对几乎每个样本，当$N$很大时，$A_N$中的元素历经$E$中各个状态，而且当$N\to\infty$时，$A_N$中元素为状态$e_i$的频率趋于$p_i$，从而对$A_N$中元素的平均$\overline{X}_N=\frac{1}{2N+1}\sum_{n=-N}^{N}X_n$趋于过程均值$m=\sum_{i=1}^{b}p_ie_i$。

关于遍历性的定义式(4.6)—(4.9)，当t只取非负实数（非负正整数）时，相应的积分和求和就限制在$[0,\infty)$上。例如，相应于(4.6)和(4.7)式的定义改为

$$\overline{X}=\lim_{T\to\infty}\frac{1}{T}\int_0^T X(t)\mathrm{d}(t)=m \tag{4.10}$$

或

$$\overline{X}=\lim_{N\to\infty}\frac{1}{N+1}\sum_{k=0}^{N}X(k)=m. \tag{4.11}$$

由遍历性的定义式，很自然要问等式左边极限是否存在？如果极限存在，因等式左边是随机变量的极限，所以要问在什么条件下它能等于常数m？关于第一个问题，主要是数学上的考虑。1931年Birkhoff已证明了只要$E|X(t)|<\infty$，则对几乎所有样本，(4.6)左边的极限一定存在.第二个问题正是我们要在本节回答的

先考虑平稳过程均值的遍历性问题。

定理 4.1 （均值遍历性定理）(i)设$X=\{X_n, n=0,\pm1,\cdots\}$为平稳序列，其协方差函数为$R(\tau)$，则$X$有遍历性的充分必要条件是

$$\lim_{N\to\infty}\frac{1}{N}\sum_{\tau=0}^{N-1}R(\tau)=0.$$

(ii) 若$X=\{X(t): -\infty<t<\infty\}$为平稳过程，则$X$有遍历性的充分必要条件是

$$\lim_{T\to\infty}\frac{1}{T}\int_0^{2T}\left(1-\frac{\tau}{2T}\right)R(\tau)\mathrm{d}\tau=0.$$

证 由于离散场合和连续时间场合证明的思路相同，所以我们仅证明连续时间的均值遍历性定理。首先计算 $\overline{X}$ 的均值和方差。记

$$\overline{X}_T = \frac{1}{2T}\int_{-T}^{T} X(t)\mathrm{d}(t),$$

则*

$$E\overline{X} = E\lim_{T\to\infty}\overline{X}_T = \lim_{T\to\infty}E\overline{X}_T = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}EX(t)\,\mathrm{d}t = m,$$

如图 4.1 $\mathrm{Var}(\overline{X}) = E\lim_{T\to\infty}\left[\frac{1}{2T}\int_{-T}^{T}(X(t)-m)\,\mathrm{d}t\right]^2$

$$= \lim_{T\to\infty}\frac{1}{4T^2}E\left[\int_{-T}^{T}(X(t)-m)\,\mathrm{d}t\right]^2$$

$$= \lim_{T\to\infty}\frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T}E[(X(t)-m)(X(s)-m)]\,\mathrm{d}t\mathrm{d}s$$

$$= \lim_{T\to\infty}\frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T}R(t-s)\,\mathrm{d}t\mathrm{d}s. \qquad (4.12)$$

在上述积分中，令 $\tau = t-s$，$v = t+s$，则变换的雅可比行列式值为
$$J = \begin{vmatrix} 1 & -1 \\ 1 & 1 \end{vmatrix}^{-1} = \frac{1}{2},$$

积分区域变换为 $-2T \leqslant \tau \pm v \leqslant 2T$（如图 4.2）。

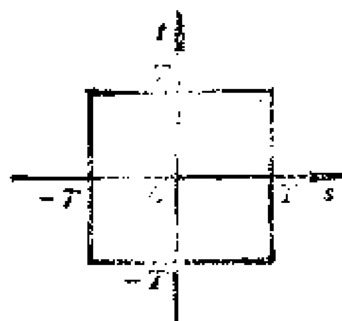


图 4.1

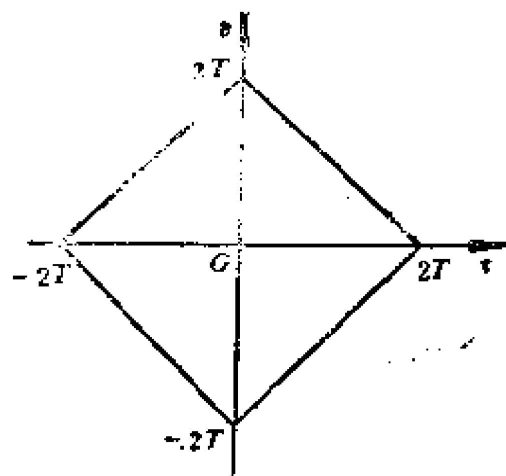


图 4.2

* 在定理条件下，所涉及运算次序的交换是允许的。

注意到 $R(\tau)$ 是偶函数，故(4.12)等于

$$\lim_{T\to\infty}\frac{1}{4T^2}\cdot\frac{1}{2}\iint_{\diamond}R(\tau)\mathrm{d}\tau\mathrm{d}v=\lim_{T\to\infty}\frac{1}{8T^2}\int_{-2T}^{2T}R(\tau)\mathrm{d}\tau\int_{-(2T-|\tau|)}^{2T-|\tau|}\mathrm{d}v$$

$$=\lim_{T\to\infty}\frac{1}{4T^2}\int_{-2T}^{2T}R(\tau)(2T-|\tau|)\mathrm{d}\tau$$

$$=\lim_{T\to\infty}\frac{1}{2T^2}\int_{0}^{2T}R(\tau)(2T-\tau)\mathrm{d}\tau$$

$$=\lim_{T\to\infty}\frac{1}{T}\int_{0}^{2T}R(\tau)\left(1-\frac{\tau}{2T}\right)\mathrm{d}\tau.$$

故关于均值的遍历性定理就化为上式极限是否趋于零的问题。由均方收敛定义知这确实是等价的，从而证明了定理。

由此定理，可以推出一些判断平稳过程均值有遍历性的充分条件：

推论 4.1 若 $\int_{-\infty}^{\infty}|R(\tau)|\mathrm{d}\tau<\infty$，则均值遍历性成立。

这是由于当 $0\leqslant\tau\leqslant 2T$ 时，$|(1-\tau/2T)R(\tau)|\leqslant|R(\tau)|$，故

$$\frac{1}{T}\left|\int_{0}^{2T}\left(1-\frac{\tau}{T}\right)R(\tau)\,\mathrm{d}\tau\right|\leqslant\frac{1}{T}\int_{0}^{2T}|R(\tau)|\mathrm{d}\tau$$

$$\leqslant\frac{1}{T}\int_{0}^{\infty}|R(\tau)|\mathrm{d}\tau\to 0$$

推论 4.2 对平稳序列而言，若 $R(\tau)\to 0$ $(\tau\to\infty)$，则均值遍历性成立。

证 因为 $R(\tau)\to 0$ $(\tau\to\infty)$，故由 Stoltz 定理知

$$\lim_{N\to\infty}\frac{1}{N}\sum_{\tau=1}^{N-1}R(\tau)=\lim_{N\to\infty}R(N-1)=0.$$

从而序列的均值有遍历性。

关于协方差函数 $R(\tau)$ 的遍历性定理，我们可以考虑随机过程 $Y_\tau=\{Y_\tau(t),-\infty<t<\infty\}$，其中

$$Y_\tau(t)=(X(t+\tau)-m)(X(t)-m),$$

则 $EY_\tau(t)=R(\tau)$。由定理 4.1 的证明过程知道，均值有遍历性等价于 $\mathrm{Var}(\overline{X})=0$。因此可以类推协方差函数 $R(\tau)$ 有遍历性等价于 $\mathrm{Var}(\overline{R}(\tau))=0$。把 $\mathrm{Var}(\overline{R}(\tau))$ 写开，我们有

定理 4.2 （协方差函数遍历性定理）设 $X=\{X(t),\ -\infty<t<\infty\}$ 为平稳过程，其均值函数为零，则协方差函数有遍历性的充分必要条件是

$$\lim_{T\to\infty}\frac{1}{T}\int_0^{2T}\left(1-\frac{\tau_1}{2T}\right)(B(\tau_1)-R^2(0))\mathrm{d}\tau_1=0,$$

其中

$$B(\tau_1)=EX(t+\tau+\tau_1)X(t+\tau_1)X(t+\tau)X(t).$$

对于定义在 $[0,\infty)$ 上的平稳过程，只要把遍历性理解为 (4.10)，(4.11) 等式。则定理 4.1 和定理 4.2 仍成立。

关于协方差函数的遍历性，由于牵涉到过程的 4 阶矩，一般很难验证。但对于 Gauss 过程来说，问题要简单得多，比如我们有如下的结果。

定理 4.3 设 $X=\{X_n, n=0,\pm 1,\cdots\}$ 是均值为 0 的 Gauss 平稳过程。如果

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}R^2(k)=0,$$

则 Gauss 过程的协方差函数有遍历性。

例 4.10 设 $X(t)=a\cos(\omega t+\Theta)$，$\Theta\sim U(0,2\pi)$，$\omega\neq 0$，则 $X=\{X(t),-\infty<t<\infty\}$ 的均值有遍历性。

证 首先

$$EX(t)=\frac{1}{2\pi}\int_0^{2\pi}a\cos(\omega t+\theta)\mathrm{d}\theta=0.$$

其协方差函数为

$$\begin{aligned}EX(t+\tau)X(t)&=Ea^2\cos(\omega(t+\tau)+\Theta)\cos(\omega t+\Theta)\\&=\frac{a^2}{2\pi}\int_0^{2\pi}\cos((t+\tau)\omega+\theta)\cos(t\omega+\theta)\mathrm{d}\theta\end{aligned}$$

$$= \frac{a^2}{4\pi}\int_0^{2\pi}[\cos((2t+\tau)\omega+2\theta)+\cos\tau\omega]\mathrm{d}\theta$$

$$=\frac{a^2}{2}\cos\tau\omega,$$

故 X 为平稳的。由于

$$\frac{1}{T}\int_0^{2T}\left(1-\frac{\tau}{2T}\right)R(\tau)\mathrm{d}\tau=\frac{a^2}{2T}\int_0^{2T}\left(1-\frac{\tau}{2T}\right)\cos\omega\tau\,\mathrm{d}\tau$$

$$=\frac{a^2}{2T}\cdot\frac{\sin 2T\omega}{\omega}-\frac{a^2}{4T^2}\int_0^{2T}\tau\cos\omega\tau\mathrm{d}\tau. \qquad (4.13)$$

但由微分第二中值定理知

$$\left|\int_0^{2T}\tau\cos\omega\tau\,\mathrm{d}\tau\right| = \left|2T\int_{\xi}^{2T}\cos\omega\tau\,\mathrm{d}\tau\right|\leqslant\frac{4T}{\omega},$$

把此结果代入(4.13)式，即知

$$\frac{1}{T}\int_0^{2T}\left(1-\frac{\tau}{2T}\right)R(\tau)\,\mathrm{d}\tau\to 0 \quad (T\to\infty),$$

故由定理 4.1 知均值遍历性成立。(对本例来说，读者可直接验证 $\frac{1}{2T}\int_{-T}^{T}X(t)\,\mathrm{d}t\to 0$)。

例 4.11 设随机变量序列 $\{X_n, n=0,\pm1,\cdots\}$ 满足

$$X_n=\sum_{k=1}^{m}(A_k\cos n\omega_k+B_k\sin n\omega_k),$$

其中 $A_1,\cdots,A_m$；B_1，…，B_m 是均值为 0 且两两不相关的随机变量，又 $EA_k^2=EB_k^2=\sigma_k^2$，$1\leqslant k\leqslant m$，$0<\omega_k<2\pi$ 试考察其均值的遍历性。

解 与例 4.6 类似，可以算得

$$EX_n=0, \quad n=0,\pm1,\cdots$$

$$R(\tau)=\sum_{k=0}^{m}\sigma_k^2\cos\tau\omega_k.$$

由三角求和公式 $\frac{1}{2}+\cos x+\cdots+\cos nx=\frac{\sin\left(n+\frac{1}{2}\right)x}{2\sin(x/2)}$，得

$$\frac{1}{N}\left|\sum_{\tau=1}^{N-1}R(\tau)\right| = \frac{1}{N}\left|\sum_{\tau=0}^{N-1}\sum_{k=0}^{m}\sigma_k^2\cos\tau\omega_k\right|$$

$$\leqslant \frac{1}{N}\sum_{k=0}^{m}\sigma_k^2\left|\sum_{\tau=0}^{N-1}\cos\tau\omega_k\right|$$

$$= \frac{1}{N}\sum_{k=0}^{m}\sigma_k^2\left|\frac{\sin(N-1/2)\omega_k}{2\sin(\omega_k/2)} - \frac{1}{2}\right| \leqslant \frac{1}{N}\sum_{k=0}^{m}\sigma_k^2$$

$$\cdot \frac{1}{2\sin(\omega_k/2)} \to 0$$

故平稳序列 X 的均值有遍历性。

例 4.12 考察例 4.7 中滑动平均序列均值的遍历性。

解 由例 4.7 知

$$R(\tau) = \begin{cases}\sigma^2(a_k a_{k-\tau} + \cdots + a_{\tau+1}a_1), & 若 \tau \leqslant k-1, \\ 0, & 若 \tau \geqslant k.\end{cases}$$

由定理 4.1 的推论 4.2 立得滑动平均过程的均值有遍历性。

例 4.13 研究例 4.8 中随机电报信号的均值有否遍历性。

解 由例 4.8 知 $R(\tau) = I^2 e^{-2\lambda|\tau|}$，其中 $\lambda>0$，由于

$$\int_{-\infty}^{\infty} R(\tau)d\tau = 2I^2\int_0^{\infty} e^{-2\lambda\tau}d\tau = \frac{I^2}{\lambda} < \infty,$$

由定理 4.1 的推论 4.1 立得随机电报信号的均值有遍历性。

在应用中，如果实验记录 $x(t)$ 只在时间区间$[0, T]$中给出，则有如下的近似估计式：

$$m \approx \hat{m}_T = \frac{1}{T}\int_0^T x(t)dt, \tag{4.14}$$

$$R(\tau) \approx \hat{R}_T(\tau) = \frac{1}{T-\tau}\int_0^{T-\tau}(x(t)-m)(x(t+\tau)-m)dt$$

$$= \frac{1}{T-\tau}\int_0^{T-\tau}x(t)x(t+\tau)dt - \hat{m}_{T-\tau}^2, \tag{4.15}$$

其中 $0\leqslant\tau\leqslant T$。但在实际中 $x(t)$的表达式是未知的。因而通常可用数值方法来计算(4.14)和(4.15)式。先把$[0, T]$等分为 N 个长为$\Delta t = T/N$的时间间隔，然后在时刻 $t_k = (k-1/2)\Delta t$, $k=1$,

$2, \cdots N$，对$X(t)$采样得到N个采样值$x_k = x(t_k)$，$k=1, 2, \cdots N$。把(4.14)中的积分用和号代替，即得

$$\hat{m}_T \approx \frac{1}{T}\sum_{k=1}^{N} x_k \Delta t = \frac{1}{N}\sum_{k=1}^{N} x_k.$$

对$\tau_r = r\Delta t$时的自相关函数可用下式来估计：

$$\hat{R}_T(\tau_r) \approx \frac{1}{T-\tau_r}\sum_{k=1}^{N-r} x_k x_{k+r}\Delta t - \left(\frac{1}{T-\tau_r}\sum_{k=1}^{N-r} x_k \Delta t\right)^2$$

$$= \frac{1}{N-r}\sum_{k=1}^{N-r} x_k x_{k+r} - \left(\frac{1}{N-r}\sum_{k=1}^{N-r} x_k\right)^2,$$

$$r = 0, 1, \cdots, \quad m < N.$$

上面两式的计算和图形容易由计算机来实现，现在的问题是Δt取多小为好？用和号近似积分，原则上Δt应该越小越好。但由于技术原因（如采样必定有一定的时间间隔）。故Δt实际上不可能取得过小。理论表明，如果$x(t)$，$0 \leqslant t \leqslant T$，的富里埃变换$H(j\omega)$有截止角频率$\omega_c$（即$H(j\omega)$在频率域$\omega \leqslant \omega_c$上非零，而$\omega > \omega_c$时$H(j\omega)=0$），则当$\Delta t \leqslant \frac{\pi}{\omega}$（或取$N > \frac{\omega_c T}{\pi}$）时，$x_k$，$k=1, 2, \cdots N$可以代表函数$x(t)$在$[0, T]$上的全部信息（这是所谓的采样定理）。这儿$j=\sqrt{-1}$表示虚数单位。

最后需要指出的是遍历性定理的条件对一般工程上遇到的平稳过程都能满足。但是真要去验证它们是否成立是很困难的。因此在应用中，一般都是先假定所遇到的平稳过程有遍历性，然后根据试验来检验这个假设是否合理。

平稳过程的遍历性在应用中有十分重要的意义。首先如本节开始所言，只要利用较长时间内的一段样本值就能估计这个过程的均值和协方差函数，其次还可直接用来求出电子系统中信号和噪声过程的直流分量，平均功率等。例如，若$X(t)$表示噪声电压（或电流），$x(t)$表示某次观察到的噪声电压（或电流）曲线。若过程X有遍历性，则由（4.10）知噪声电压（或电流）的直流分

量为

$$m=\overline{m}=\lim_{T\to\infty}\frac{1}{T}\int_0^T x(t)\,dt.$$

这个直流分量很容易用实验方法确定。又可算出

$$\sigma^2=R(0)=\lim_{T\to\infty}\frac{1}{T}\int_0^T (x(t)-m)^2 dt.$$

不难看出，$R(0)$恰为噪声电压(或电流)消耗在 1 欧姆 电 阻上的交流功率，而 σ 则代表电压(或电流)的有效值。

4.3 平稳过程的协方差函数和功率谱密度

如同信号中的确定性函数一样，对平稳过程，我们既可以在时间域上来研究其性质，也可以利用 Fourier 变换在频率域上研究。具体在哪个域内研究要视问题而定。本节我们主要讨论平稳过程协方差函数及其对应频率域上的频率结构。以下$X=\{X(t),t\in T\}$均表示实或复值宽平稳过程。

4.3.1 协方差函数

对平稳过程 X 的协方差函数 $R(\tau)$，容易由定义得到如下性质：

1. 对称性，即 $R(-\tau)=R(\tau)$。
2. 有界性，即 $|R(\tau)|\leqslant|R(0)|$。

此外，可以推得 $R(\tau)$ 有非负定性这一重要性质。即对任意的时刻 t_n 及实数 a_n，$n=1,2,\cdots,N$。有

3. $\sum_{n=1}^{N}\sum_{m=1}^{N}a_n a_m R(t_n-t_m)\geqslant 0$。

关于平稳过程导数的协方差，我们有如下结果：

4. $\mathrm{Cov}(X^{(n)}(t),X^{(n)}(t+\tau))=(-1)^n R^{(2n)}(\tau)$。　　(4.16)

证　我们只对 $n=1$证明，其余情况不难用归纳法证明。同时假定下面所涉及的运算可以交换次序以及所涉及的导数都存在。

由 $EX'(t)=(EX(t))'=0$，因此

$$
\begin{aligned}
&\mathrm{Cov}(X'(t),X'(t+\tau))=EX'(t)X'(t+\tau)\\
&=E\left(\lim_{\Delta t\to 0}\frac{X(t+\Delta t)-X(t)}{\Delta t}\lim_{\Delta\tau\to 0}\frac{X(t+\tau+\Delta\tau)-X(t+\tau)}{\Delta\tau}\right)\\
&=\lim_{\Delta t\to 0}\lim_{\Delta\tau\to 0}\frac{1}{\Delta t\Delta\tau}E(X(t+\Delta t)-X(t))(X(t+\tau+\Delta\tau)\\
&\quad -X(t+\tau))\\
&=\lim_{\Delta t\to 0}\lim_{\Delta\tau\to 0}\frac{1}{\Delta t\Delta\tau}(R(\tau+\Delta\tau-\Delta t)-R(\tau-\Delta t)\\
&\quad -R(\tau+\Delta\tau)+R(\tau))\\
&=\lim_{\Delta t\to 0}\frac{1}{\Delta t}(R'(\tau-\Delta t)-R'(\tau))=-R''(\tau).
\end{aligned}
\tag{4.17}
$$

表 4.1 中列举了工程中经常遇到的一些协方差函数。下面介绍其中几个。

A.

$$
R(\tau)=\begin{cases}1-\dfrac{|\tau|}{T}, & |\tau|\leqslant T,\\ 0, & |\tau|>T.\end{cases}
$$

$R(\tau)$ 的图形如表 4.1. 1 中所示的三角形。$R(\tau)$ 在 $\tau=0$ 连续，但不可微。由性质 4 知 $X'(t)$ 不存在。

B. $R(\tau)=\mathrm{e}^{-k|\tau|}$,　　$k>0$.

其图形如表 4.1. 2 中所示，为指数衰减型。这是最常用的协方差函数。图形在 τ 轴上方且为凸的(即二阶导数为正)。在 $\tau=0$ 连续但不可微，故 $X'(t)$ 也不存在。

C.　$R(\tau)=(1-k|\tau|)\mathrm{e}^{-k|\tau|}$, $k>0$.

其图形如表 4.1. 4 所示。由图知当 $|\tau|>1/k$ 时，图形在 τ 轴下方，且在 $|\tau|=2/k$ 时达到最小值。由于 $R(\tau)$ 在 $\tau=0$ 导数不存在，故 $X'(t)$ 不存在。

D. $R(\tau)=\exp\left\{-\dfrac{k^2}{2}\tau^2\right\}$

该图与均值为零，方差为 $1/k^2$ 的正态密度函数仅差一个常数因子。$\tau=0$ 时 $R(\tau)$ 任意次可微，故 $X^{(n)}(t)$ 存在。图见表 4.1. 5.

表 4.1

	$R(\tau)$	$S(\omega)$
1	1; $-T$, O, T, τ	T; $\frac{4\sin^2(\omega T/2)}{T\omega^2}$; O, $2\pi/T$, ω
2	1; $e^{-a\lvert\tau\rvert}$; O, τ	$2/a$; $\frac{2a}{a^2+\omega^2}$; O, ω
3	1; $(1+k\lvert\tau\rvert)e^{-k\lvert\tau\rvert}$; O, τ	$2/k$; $\frac{2k^3}{(k^2+\omega^2)^2}$; O, ω
4	1; $(1-k\lvert\tau\rvert)e^{-k\lvert\tau\rvert}$; O, τ	$\frac{2k\omega^2}{(k^2+\omega^2)^2}$; O, ω
5	1; $e^{-k\tau^2/2}$; O, τ	$1/\sqrt{k}$; $\frac{1}{\sqrt{k}}e^{-\omega^2/2k}$; O, ω
6	$e^{-a\lvert\tau\rvert}\cos\omega_0\tau$; O, τ	$1/a$; $-\omega_0$, O, ω_0, ω
7	$\frac{\omega_0}{\pi}$; $\frac{\pi}{\omega_0}$; $\frac{\sin\omega_0\tau}{\pi\tau}$; O, τ	1; $-\omega_0$, O, ω_0, ω
8	1; O, τ	2π; O, ω
9	1; O, τ	1; O, ω

续表 4.1

	$R(\tau)$	$S(\omega)$
10	$\cos\omega_0\tau$ O τ	π $-\omega_0$ O ω_0 ω

4.3.2 几个常见随机信号的协方差函数

1. 振幅调制波

设$\{Y(t),t\in R\}$是一个实平稳过程。λ_0 为复数，则

$$Z(t)=Y(t)e^{j\lambda_0 t},\ t\in R. \tag{4.18}$$

是一个复值的平稳过程。λ_0 为角频率。$Z(t)$ 称为由 $Y(t)$确定的振幅调制信号。分别以 R_Z和 R_Y 表示过程 Z 和 Y 的协方差函数，则

$$R_Z(\tau)=R_Y(\tau)e^{j\lambda_0\tau}.$$

特别若 $R_Y(\tau)=e^{-k|\tau|}$，则

$$R_Z(\tau)=e^{-k|\tau|+j\lambda_0\tau}.$$

这儿及下面 j 均表示虚数单位，即 $j=\sqrt{-1}$。

实数形式的振幅调制波表示为

$$X(t)=Y(t)\cos(\omega t+\Theta),\ t\in R,$$

其中随机相位Θ为$(0,\ 2\pi)$中的均匀分布，且和随机振幅 $Y(t)$ 独立，则其协方差函数当 $EY(t)=0$时 为

$$R_X(\tau)=R_Y(\tau)\cdot\frac{1}{2\pi}\int_0^{2\pi}\cos(\omega t+\theta)\cos(\omega(t+\tau)+\theta)\mathrm{d}\theta$$

$$=\frac{1}{2}R_Y(\tau)\cos\omega\tau.$$

若 $R_Y(\tau)=e^{-k|\tau|}$，则

$$R_X(\tau)=e^{-k|\tau|}\cos\omega\tau.$$

2. 频率调制波

为简单起见，仅讨论由平稳过程$Y=\{Y(t),t\in R\}$确定的相位调制信号：

$$X(t)=\cos Y(t),\ t\in R$$

设 Y 为零均值Gauss过程。我们来求过程 X 的协方差函数。令 $C=\sqrt{R_Y(0)}$，则

$$EX(t)=E\cos Y(t)=E\frac{1}{2}(e^{jY(t)}+e^{-jY(t)})$$

设$Y(t)$的矩母函数为 $g(s)$，则 $g(s)=Ee^{sY(t)}=\exp\{C^2s^2/2\}$，以 $s=j$ 形式地代入上式(这时矩母函数称为特征函数)即有

$$E\cos Y(t)=\frac{1}{2}\left(\exp\left\{\frac{-C^2}{2}\right\}+\exp\left\{-\frac{C^2}{2}\right\}\right)=\exp\left\{-\frac{C^2}{2}\right\}$$
$$=\exp\{-R_Y(0)/2\}.$$

由三角公式

$$\cos\alpha\cos\beta=\frac{1}{4}(e^{j(\alpha+\beta)}+e^{-j(\alpha+\beta)}+e^{j(\alpha-\beta)}+e^{-j(\alpha-\beta)})$$

经过计算后也可以得到

$$EX(t)X(t+\tau)=e^{-R_Y(0)}\mathrm{ch}(R_Y(\tau)),$$

因此我们得到

$$R_X(\tau)=EX(t)X(t+\tau)-(EX(t))^2$$
$$=e^{-R_Y(0)}(\mathrm{ch}(R_Y(\tau))-1). \tag{4.19}$$

如果 $\tau\to\infty$ 时有 $R_Y(\tau)\to 0$，则由(4.19)式知 $R_X(\tau)\to 0$。由协方差函数性质4知 $-R''_Y(\tau)=E(Y'(t))^2>0$，故 $R_Y(\tau)$ 在 $\tau=0$ 附近有如下近似公式：

$$R_Y(\tau)\approx R_Y(0)-\frac{k^2}{2}\tau^2.$$

上式中没有一次项是因为 $R_Y(\tau)$是偶函数，代入(4.19)式，得

$$R_X(\tau)\approx\exp\{-R_Y(0)+R_Y(\tau)\}=\exp\left\{-\frac{1}{2}k^2\tau^2\right\}$$

即近似为正态密度函数型的协方差函数。

3. 平方检波

设$\{Y(t), t\in R\}$为零均值平稳 Gauss 过程。令

$$X(t)=Y^2(t),\ t\in R.$$

通过计算可得过程 X 和 Y 的协方差函数有如下关系：

$$R_X(\tau)=2R_Y^2(\tau).$$

4.3.3 功率谱密度

本小节主要研究协方差函数的频率结构。我们先从确定性时间函数的能量，能谱密度，功率和功率谱等概念出发，然后引入平稳过程功率谱的概念。

设$x(t)$是实轴上以$2T$为周期的函数，在$[-T, T]$上只有有限个第一类间断点(也可改为其他条件)，则$x(t)$有富里埃展开，即

$$x(t)=\sum_{n=-\infty}^{\infty}A(n)e^{-jn\omega t},$$

其中 $\omega=\pi/T$.

$$A(n)=\frac{1}{2T}\int_{-T}^{T}x(t)e^{jn\omega t}dt.$$

$A(n)$ 一般为复数，由定义显然有$A(n)$的共轭 $\overline{A}(n)=A(-n)$。$\frac{1}{2}A(0)$称为直流分量，$|A(1)|$称为基波ω的振幅，$|A(n)|$称为谐波$n\omega$的振幅。现考虑$x(t)$按频率在$[-T,T]$上能量的分解。(为了便于理解诸物理术语，可以把$x(t)$设想为加在1欧姆电阻上的电压)。若

$$\int_{-T}^{T}x^2(t)dt<\infty,$$

则成立 Parsval 等式：

$$\int_{-T}^{T}x^2(t)dt=2T\sum_{n=-\infty}^{\infty}|A(n)|^2.$$

上式左边表示$x(t)$在$[-T, T]$上的总能量，或者说$x(t)$的功率为

$$\frac{1}{2T}\int_{-T}^{T}x^2(t)\mathrm{d}t=\sum_{n=-\infty}^{\infty}|A(n)|^2.$$

我们称 $\omega_n=n\omega$ 为各谐波的角频率。$\lambda=n/2T$ 称为相应谐波的线频率。下图为 $x(t)$ 的线功率谱图。每条线段称为一条谱线，高度为 $|A(n)|^2$，$|A(-n)|^2+|A(n)|^2$ 为基波或第 $(n-1)$ 个谐波的功率。

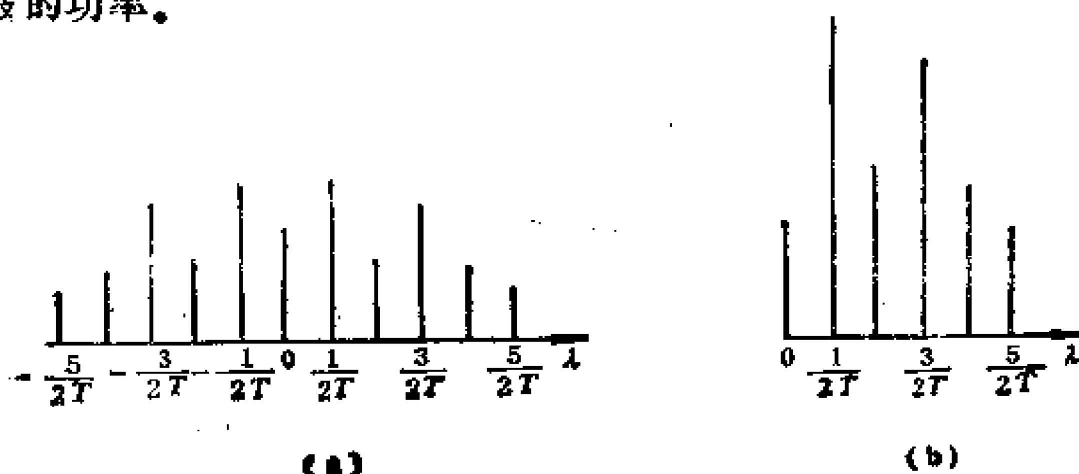


图 4.3　周期为 $2T$ 的函数功率谱

图 4.3(b) 是把功率谱都集中在右半直线上，用它表示的称为半功率谱。这儿"半"字不是量值的一半，而是区域的一半。因为除 0 点外，半功率谱的值为功率谱在同处的二倍。由于上述谱线是离散的，所以我们把这种功率谱称为离散谱。上面的分析告诉我们，如果 $[-T, T]$ 上 $x(t)$ 的总能量有限，则总功率可分解为各谐波的功率之和。如果 $x(t)$ 不是周期的，但总能量有限，即

$$\int_{-\infty}^{\infty}x^2(t)\mathrm{d}t<\infty, \tag{4.20}$$

则 $x(t)$ 的Fourier 变换存在或者说有频谱

$$F(\omega)=\int_{-\infty}^{\infty}x(t)\mathrm{e}^{-j\omega t}\mathrm{d}t.$$

一般 $F(\omega)$ 为复数，且 $\bar{F}(\omega)=F(-\omega)$。而 Parsval 等式仍成立：

$$\int_{-\infty}^{\infty}x^2(t)\mathrm{d}t=\frac{1}{2\pi}\int_{-\infty}^{\infty}|F(\omega)|^2\mathrm{d}\omega.$$

这公式的意义是 $x(t)$的全部能量可以按所有频率进行分解。$|F(\omega)|^2$ 称为能量谱密度。满足(4.20)式的信号称为能量型信号。

但在一般情况下，信号不满足(4.20)式，即总能量为无限，但是平均功率

$$P=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x^2(t)\,\mathrm{d}t<\infty.$$

这种信号称为功率型信号。正弦型信号、平稳过程的样本函数等时间函数，由于 t 趋于正负无穷大时都不收敛于 0，所以都不是能量型信号。但显然正弦型信号是功率型的。为了利用Fourier变换给出平均功率的谱表达式，我们可以按如下方式来考虑。令

$$x_T(t)=\begin{cases}x(t), & |t|\leqslant T,\\ 0, & |t|>T,\end{cases} \tag{4.21}$$

则 $x(t)$ 在有限区间上的能量总是有限的。故 $x_T(t)$ 的Fourier 变换存在。

$$F(\omega,T)=\int_{-\infty}^{\infty}x_T(t)\mathrm{e}^{-j\omega t}\mathrm{d}t, \tag{4.22}$$

且成立 Parsval 等式：

$$\int_{-\infty}^{\infty}x_T^2(t)\,\mathrm{d}t=\frac{1}{2\pi}\int_{-\infty}^{\infty}|F(\omega,T)|^2\mathrm{d}\omega. \tag{4.23}$$

两边除以 $2T$，左边即为 $x_T(t)$的平均功率。令 $T\to\infty$，则 $x(t)$在实数轴上的平均功率可表为

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-\infty}^{\infty}x_T^2(t)\,\mathrm{d}t=\frac{1}{2\pi}\lim_{T\to\infty}\int_{-\infty}^{\infty}\frac{1}{2T}|F(\omega,T)|^2\mathrm{d}\omega. \tag{4.24}$$

如果

$$\lim_{T\to\infty}\frac{1}{2T}|F(\omega,T)|^2=S(\omega)$$

存在，则(4.24)中的极限号和积分号可以交换。注意到(4.21)式，有

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x^2(t)\mathrm{d}t=\frac{1}{2\pi}\int_{-\infty}^{\infty}S(\omega)\mathrm{d}\omega. \tag{4.25}$$

我们称 $S(\omega)$ 为 $x(t)$ 的平均功率谱密度。$S(\omega)\Delta\omega$ 表示 $x(t)$ 的频率在 $[\omega,\omega+\Delta\omega]$ 中的成分对 $x(t)$ 总功率的贡献。

下面我们把平均功率谱密度的概念推广到平稳过程 $X=\{X(t),-\infty<t<\infty\}$。为此，对每条样本轨道，根据(4.22)和(4.25)(设 $S(\omega)$ 存在)，则有

$$F(\omega,T)=\int_{-T}^{T}X(t)\mathrm{e}^{-j\omega t}\mathrm{d}t \tag{4.26}$$

和

$$\frac{1}{2T}\int_{-T}^{T}X^2(t)\mathrm{d}t=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{1}{2T}|F(\omega,T)|^2\mathrm{d}\omega. \tag{4.27}$$

注意，(4.26)和(4.27)两式中的量都是随机变量，所以有意义的是它们的平均值。我们把过程 X 的平均功率谱密度定义为(如果极限存在的话)

$$S(\omega)=\lim_{T\to\infty}E\frac{1}{2T}|F(\omega,T)|^2, \tag{4.28}$$

而

$$\lim_{T\to\infty}E\left(\frac{1}{2T}\int_{-T}^{T}X^2(t)\mathrm{d}t\right) \tag{4.29}$$

定义为平稳过程 X 的平均功率。

设 $EX(t)=0$，在(4.29)中交换积分和期望运算次序，则

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}EX^2(t)\mathrm{d}t=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}R(0)\mathrm{d}t=R(0), \tag{4.30}$$

即 $R(0)$ 就表示过程的平均功率。在(4.28)中交换运算次序，得

$$\lim_{T\to\infty}E\int_{-\infty}^{\infty}\frac{1}{2T}|F(\omega,T)|^2\mathrm{d}\omega=\int_{-\infty}^{\infty}\lim_{T\to\infty}E\frac{1}{2T}|F(\omega,T)|^2\mathrm{d}\omega$$

$$=\int_{-\infty}^{\infty}S(\omega)\mathrm{d}\omega. \tag{4.31}$$

在(4.27)式两边取期望，然后令 $T\to\infty$，则由(4.30)和(4.31)

得

$$R(0)=\frac{1}{2\pi}\int_{-\infty}^{\infty}S(\omega)\mathrm{d}\omega. \tag{4.32}$$

(4.32)式称为平稳过程的平均功率的谱表示式。当然，这里还差一个所述运算次序能否交换的问题。可以证明，如果平稳过程的协方差函数 $R(\tau)$ 满足 $\int|R(\tau)|\mathrm{d}\tau<\infty$，则以上各种运算次序的交换都是合法的。

如果把 $-\omega$ 处的谱密度加到 ω 处，使谱密度只在正半实轴上有定义，则称之为半(功率)谱密度或单边谱密度。如用 $G(\omega)$ 表示半谱密度，则

$$G(\omega)=\begin{cases}\lim\limits_{T\to\infty}\dfrac{1}{T}E|F(\omega,T)|^2, & \omega\geqslant 0,\\ 0, & \omega<0.\end{cases}$$

$$=\begin{cases}2S(\omega), & \omega\geqslant 0,\\ 0. & \omega<0.\end{cases} \tag{4.33}$$

如图 4.4 所示。半谱密度在工程上更经常使用。

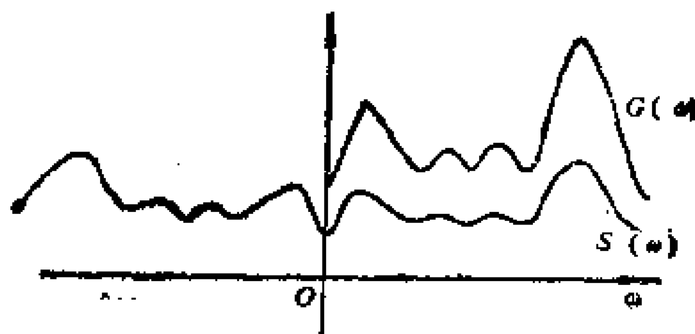


图 4.4 谱密度与半谱密度

由于谱密度在谱分析中的重要性，有必要仔细讨论它的性质，特别是它与协方差函数之间的关系。

由 $S(\omega)$ 的定义可知

$\overline{S}(\omega)=S(\omega)\geqslant 0$,

$S(-\omega)=S(\omega)$.

这是因为 $|F(\omega,T)|^2=F(\omega,T)F(-\omega,T)$ 为实的，非负偶函数。

其次，$S(\omega)$和协方差函数 $R(\tau)$是一对富里埃变换。更具体地，我们有

定理 4.4 （Wiener-Khintchine公式）若$\int|R(\tau)|d\tau<\infty$，则

$$S(\omega)=\int R(\tau)e^{-j\omega\tau}d\tau, \tag{4.34}$$

$$R(\tau)=\frac{1}{2\pi}\int S(\omega)e^{j\omega\tau}d\omega. \tag{4.35}$$

证 为避免演算的复杂性，我们仅对过程均值为 0 的情况加以证明，此时 $R(t-s)=EX(t)X(s)$。由$S(\omega)$ 的定义，有

$$\begin{aligned}S(\omega)&=\lim_{T\to\infty}\frac{1}{2T}E\int_{-T}^{T}X(t)e^{-j\omega t}dt\int_{-T}^{T}X(s)e^{j\omega s}ds\\&=\lim_{T\to\infty}\frac{1}{2T}\iint_D E(X(t)X(s))e^{-j\omega(t-s)}dtds\\&=\lim_{T\to\infty}\frac{1}{2T}\iint_D R(t-s)e^{-j\omega(t-s)}dtds,\end{aligned}$$

这儿 $D=\{-T\leqslant t\leqslant T,-T\leqslant s\leqslant T\}$。在上式积分中作变量代换：$u=t+s$，$\tau=t-s$，则积分区域 D 变换为 $D_1=\{-2T\leqslant u+\tau\leqslant 2T,-2T\leqslant u-\tau\leqslant 2T\}$（如同图 4.2）则类似于定理 4.1 的证明上式为

$$\begin{aligned}S(\omega)&=\lim_{T\to\infty}\frac{1}{4T}\iint_{D_1}R(\tau)e^{-j\omega\tau}dud\tau\\&=\lim_{T\to\infty}\int_{-\infty}^{\infty}R^T(\tau)e^{-j\omega\tau}d\tau,\end{aligned}$$

式中

$$R^T(\tau)=\begin{cases}\left(1-\frac{|\tau|}{2T}\right)R(\tau), & |\tau|\leqslant 2T,\\ 0, & |\tau|>2T.\end{cases}$$

由于 $T\to\infty$ 时，$R^T(\tau)\to R(\tau)$，故由条件 $\int|R(\tau)|d\tau<\infty$ 知极限号和积分号可交换，即有

$$S(\omega)=\int_{-\infty}^{\infty}\lim_{T\to\infty}R^T(\tau)e^{-j\omega\tau}d\tau=\int_{-\infty}^{\infty}R(\tau)e^{-j\omega\tau}d\tau.$$

由 Fourier 变换可知，$R(\tau)$ 是 $S(\omega)$ 的反 Fourier 变换，故 (4.34) 和 (4.35) 两式成立。

注。由于 $R(\tau)$ 和 $S(\omega)$ 都是偶函数，故 Wiener-Khintchine 公式还可以写为偶 Fourier 变换形式：

$$S(\omega)=2\int_0^{\infty}R(\tau)\cos\omega\tau\,d\tau, \tag{4.36}$$

$$R(\tau)=\frac{1}{\pi}\int_0^{\infty}S(\omega)\cos\omega\tau\,d\omega. \tag{4.37}$$

对平稳序列来说，设 $R(\tau),\tau=0,\pm1,\pm2,\cdots$ 为其协方差函数，且 $\sum|R(\tau)|<\infty$，$S(\omega)$ 为对应的谱密度（即在 (4.26) 中把积分号改为求和号，然后由 (4.28) 得到的平均功率谱密度），则对应于 (4.36) 和 (4.37) 的 Wiener-Khintchine 公式为

$$S(\omega)=\sum_{\tau=-\infty}^{\infty}e^{-j\omega\tau}R(\tau), \tag{4.38}$$

$$R(\tau)=\frac{1}{2\pi}\int_{-\pi}^{\pi}S(\omega)\cos\tau\omega\,d\omega. \tag{4.39}$$

表 4.1 列出了若干自相关函数及对应的谱密度。

最常见的谱密度是有理谱密度，即 $S(\omega)$ 为两个 ω 的多项式的比：

$$S(\omega)=\frac{P(\omega)}{Q(\omega)}.$$

由谱密度是 ω 的非负实值偶函数知，$S(\omega)$ 形如

$$s_0\frac{\omega^{2n}+a_{2n-2}\omega^{2n-2}+\cdots+a_2\omega^2+a_0}{\omega^{2m}+b_{2m-2}\omega^{2m-2}+\cdots+b_2\omega^2+b_0},$$

式中 $s_0\neq0$。又由于 $R(0)>0$，故 $S(\omega)$ 应在 $[0,\infty)$ 上可积，从而 $S(\omega)$ 的分母不能有实根，分母多项式次数至少应比分子高 2 以及 $s_0>0$。

例 4.14 已知谱密度为

$$S(\omega)=\frac{\omega^2+2}{\omega^4+5\omega^2+4},$$

求平稳过程 X 的协方差函数和方差。

解 由公式(4.35)，

$$R(\tau)=\frac{1}{2\pi}\int\frac{\omega^2+2}{\omega^4+5\omega^2+4}e^{j\omega\tau}d\omega.$$

由留数定理(对 $\tau\geqslant 0$ 利用上半平面围道，$\tau<0$ 时用下半平面上的围道)可算得

$$R(\tau)=\frac{1}{2\pi}2\pi j\left\{\frac{z^2+2}{z^4+5z^2+4}e^{jz|\tau|}\text{在 } z=j,2j \text{ 处的留数和}\right\}$$

$$=\frac{1}{6}(e^{-|\tau|}+e^{-2|\tau|}).$$

过程的方差为

$$R(0)=\frac{1}{6}(1+1)=\frac{1}{3}.$$

例 4.15 已知平稳序列协方差函数

$$R(\tau)=\frac{1}{2\pi}\rho^{|\tau|},\ \tau=0,\pm1,\cdots,\ |\rho|<1.$$

求谱密度函数。

解 由公式(4.38)，

$$S(\omega)=\frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty}e^{-j\omega\tau}\rho^{|\tau|}$$

$$=\frac{1}{2\pi}\left[\sum_{\tau=0}^{\infty}(\rho e^{-j\omega})^\tau+\sum_{\tau=0}^{\infty}(\rho e^{j\omega})^\tau-1\right]$$

$$=\frac{1}{2\pi}\left[\frac{1}{1-\rho e^{-j\omega}}+\frac{1}{1-\rho e^{j\omega}}-1\right]$$

$$=\frac{1}{2\pi}\cdot\frac{1-\rho^2}{|1-\rho e^{j\omega}|^2}=\frac{1}{2\pi}\cdot\frac{1-\rho^2}{1-2\rho\cos\omega+\rho^2},$$

例 4.16 如果上例中的 $R(\tau)$ 为平稳过程的协方差函数，求对应的谱密度函数。

解 由公式(4.34)，

$$S(\omega)=\int\frac{1}{2\pi}\rho^{|\tau|}e^{-j\omega\tau}d\tau$$

$$=\frac{1}{2\pi}\left[\int_0^{\infty}(\rho e^{-j\omega})^{\tau}d\tau+\int_0^{\infty}(\rho e^{j\omega})^{\tau}d\tau\right]$$

$$=\frac{1}{2\pi}\left(\frac{1}{\rho-j\omega}+\frac{1}{\rho+j\omega}\right)=\frac{\rho}{\pi(\rho^2+\omega^2)}.$$

关于由 $R(\tau)$ 求谱密度 $S(\omega)$ 的方法，常常可以利用 Fourier 变换的性质来做，比如可以利用卷积性质，微分性质等等，我们在此不再一一列举了。

最后需要指出的是在实际问题中会遇到这样一些平稳过程，它们的协方差函数或谱密度函数在常义下的 Fourier 变换不存在，比如常数或正弦型函数。但由 Fourier 分析中知道，如果允许谱密度和协方差函数中含有 δ 函数，则在广义 Fourier 变换下，仍成立 Wiener-Khintchine 公式。这主要是利用 δ 函数的如下基本性质：对任一连续函数 $f(\tau)$，

$$\int\delta(\tau-\tau_0)f(\tau)d\tau=f(\tau_0).$$

由此可得

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{j\omega\tau}d\omega=\delta(\tau), \tag{4.40}$$

$$\int\delta(\tau)e^{-j\omega\tau}d\tau=1. \tag{4.41}$$

由(4.40)，当协方差函数为常数 1 时，对应的谱密度为 $2\pi\delta(\omega)$。反之，当谱密度为 1 时，由(4.40)知协方差函数为 $\delta(\tau)$。再由

$$\cos\omega\tau=\frac{1}{2}(e^{j\omega\tau}+e^{-j\omega\tau}),\quad \sin\omega\tau=\frac{1}{2j}(e^{j\omega\tau}-e^{-j\omega\tau}),$$

可以得到当谱密度为 $a\cos\omega\tau_0$ 时，其对应的协方差函数为

$$R(\tau)=\frac{1}{2\pi}\int a\cos\omega\tau_0 e^{j\omega\tau}d\omega$$

$$= \frac{a}{4\pi}\left[\int e^{j\omega(\tau+\tau_0)}d\omega + \int e^{j\omega(\tau-\tau_0)}d\omega\right]$$

$$= \frac{a}{2}(\delta(\tau+\tau_0)+\delta(\tau-\tau_0)).$$

表 4.1 中的 8—10 用图表达了上述关系。

谱密度为常数 S_0 的平稳过程称为白噪声过程。它是白光的谱，所以称之为白噪声。白噪声的协方差函数为脉冲函数：

$$R(\tau) = \frac{1}{2\pi}\int S_0 e^{j\omega\tau}d\omega = S_0\delta(\tau).$$

故在白噪声过程中，不同时刻之间是不相关的。容易看出（定理 4.1 的推论 1），白噪声过程是遍历的。白噪声是一种理想化的数学模型，它的平均功率 $R(0)$ 是无限的，故在物理上是不能实现的。实用频谱分析中，如果某种噪声在比实际考虑的有用频带宽得多的范围内具有较平坦的谱密度时，可以把它近似当作白噪声来处理。例如电子管中散弹噪声和电子设备中的热噪声等。

对平稳序列 $X=\{X(n), n=0, \pm1, \pm2, \cdots\}$来说，如果谱密度为常数 S_0，则称为白噪声序列，其协方差函数为

$$R(\tau) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_0 \cos\tau\omega d\omega$$

$$= \begin{cases} S_0, & \tau = 0, \\ 0, & \tau = \pm1, \ \pm2, \cdots. \end{cases}$$

即不同时刻之间不相关。如果进一步假定 $X(n)$为正态分布，则称为高斯白噪声序列。其方差为 S_0。这时由多维正态分布性质知$\{X(n), n=\cdots, -1, 0, 1, \cdots\}$是相互独立的。

例 4.17 设信号输入 $\{\varepsilon_n, n=0, 1, \cdots\}$ 为一列零均值高斯白噪声，求输出 X_n

$$X_n = a_1\varepsilon_n + a_2\varepsilon_{n-1}, \quad n=1, 2, \cdots$$

的谱密度。

解 由例 4.7 知$\{X_n, n=1,2,\cdots\}$为平稳序列且其协方差函数$R(\tau)$为

$$R(\tau)=\begin{cases}a_1^2+a_2^2, & \tau=0,\\ a_1a_2, & \tau=-1,1,\\ 0, & 其他。\end{cases}$$

故对应的谱密度由(4.38)知为

$$S(\omega)=\sum_{-\infty}^{\infty}R(\tau)e^{-j\omega\tau}=(a_1^2+a_2^2)+a_1a_2(e^{-j\omega}+e^{j\omega})$$

$$=a_1^2+2a_1a_2\cos\omega+a_2^2.$$

这是一个周期为2π的周期函数。

例 4.18 信号分析中，常用复数形式来表示信号。所以考虑复值平稳过程是有意义的。设平稳过程X为

$$X(t)=a\,e^{j(\omega t-\varphi)},$$

其中W是有概率密度$f_W(\omega)$的随机变量，a为常数。容易验证，其协方差函数$R(\tau)$为

$$R(\tau)=a^2\int f_W(\omega)e^{j\omega\tau}d\omega.$$

由 Wiener-Khintchine 公式知其谱密度为

$$S(\omega)=2\pi a^2f_W(\omega).$$

这一公式在物理中有一有趣的应用。考虑 Doppler 效应。设一谐波振荡器位于x轴的点P上。它以速度v沿x轴方向移动。发射信号为$e^{j\omega_0 t}$. 位于原点O的接受器接受到的信号为

$$s(t)=a\,e^{j\omega_0(t-r/c)}$$

这儿$r=r_0+vt$是从O到P的距离，c是传播速度，则$s(t)$可写为

$$s(t)=a\,e^{j(\Omega t-\varphi)},\ \Omega=\omega_0(1-v/c),\ \varphi=r_0\omega_0/c.$$

若以$f_v(v)$表示v的概率密度。由W与v的线性关系$W=\omega_0(1-v/c)$即知W的概率密度函数为

$$f_W(\omega)=\frac{c}{\omega_0}f_v\left[\left(1-\frac{\omega}{\omega_0}\right)c\right].$$

因此接受信号的谱等于

$$S(\omega)=\frac{2\pi a^2c}{\omega_0}f_v\left[\left(1-\frac{\omega}{\omega_0}\right)c\right].$$

如果振荡器沿与 x 轴成一定角度的方向移动，上式仍成立，只不过此时 v 表示速度在 x 轴上的分量。

注意到若 $v=0$，则 $S(\omega)=2\pi a^2\delta(\omega-\omega_0)$，而在原点接受到的信号 $s(t)$ 由于运动而使它的谱密度变宽，如图 4.5 所示。

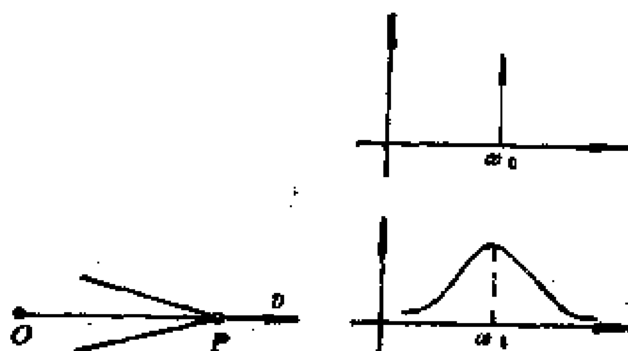


图 4.5

Wiener-khintchine 公式本质上是关于平稳过程协方差函数的 Fourier 变换问题。由这自然会想到平稳过程本身是否可以作 Fourier 变换？例 4.6 说明只要随机振幅两两不相关，则它们的三角和是平稳的。反过来，如果平稳过程 X 有周期 T，则它可展成 Fourier 级数：

$$X(t)=\sum_{-\infty}^{\infty}c_n e^{jn\omega_0 t},$$

这儿

$$c_n=\frac{1}{T}\int_{-T/2}^{T/2}X(t)e^{-jn\omega_0 t}dt,\ \omega_0=2\pi/T.$$

且振幅 $\{c_n, n=\cdots,-1,0,1\}$ 是一列两两不相关的随机变量。如果平稳过程 X 没有周期，仍有某种形式的 Fourier变换，但需要有更多的数学工具才能处理这一问题。我们不再详细讨论了。

4.4 平稳序列的预报

平稳过程研究的另一重要课题是预报问题。即在已知时刻 t 之前过程取值的条件下，如何来预报（又称预测）过程在将来的值。预报问题在实际问题中有非常广泛的应用。例如铁道部门经常要预报下一个春节期间的客流量以便安排好春运工作，气象部门要预报降水量以便更好地安排各地的工农业生产，工业自动生产线上要利用预报来进行自动控制，市场要预报产品的需求量……。为了避免理论上不必要的麻烦，讲清如何预报，我们仅处理平稳序列的预报问题。

4.4.1 一般预报理论

设 X 表示某次"试验"的一个结果。假定 X 是未知的，在将来才会被观察到。现在要求预报 X 的值。设 $\hat{X}$ 是 X 的一个预报，用 $X-\hat{X}$ 表示预报误差。要求在某种意义下预报误差越小越好。那末什么是"小"呢？一般 $X-\hat{X}$ 是随机变量，因此必须在平均的意义下使 $X-\hat{X}$ 或 $X-\hat{X}$ 的某个函数越小越好。不管我们如何选择误差函数，一个合理的假设是当误差为零时它应最小而当误差偏离零时它应随之增加。这儿我们采用均方误差 $E(X-\hat{X})^2$ 来衡量预报误差，即在均方意义下，使预报误差达到最小。采用均方误差主要有二个原因。一是如果预报均方误差越小，则预报越正确，反之则预报可能会产生较大的误差；二是在数学上便于处理，在此准则下我们的预测理论能在许多特别有用的场合中导出最佳预值的明确的表达式。

如果有估计 $\hat{X}$，使 $E(X-\hat{X})^2$ 达到最小，则称 $\hat{X}$ 为 X 的最小均方误差预报。

我们的问题是找出这样一个预报，使在一切允许的预报中有最小的均方误差。为此，我们必须指明预报 $\hat{X}$ 所能取值的范围。

一般 $\hat{X}$ 的形式由试验的知识或可能的 X 的分布函数的知识来确定。以下用 $\mathscr{H}$ 表示允许的预报类。

例 4.19 设某次试验的一个结果为 X，其均值为 μ，方差为 σ^2，此外对 X 没有进一步的信息。则在均方误差最小的意义下对 X 的最好预报是 $\hat{X}=\mu$。由于 X 没有进一步的信息可利用，故把预报空间 $\mathscr{H}$ 取为所有的实数集合。我们可以选取任一实数 a 作为对 X 的预报。即 $\hat{X}=a$，这时

$$
\begin{aligned}
E(X-\hat{X})^2 &= E[(X-\mu)+(\mu-\hat{X})]^2 \\
&= E(X-\mu)^2+2E(X-\mu)(\mu-\hat{X})+E(\mu-\hat{X})^2.
\end{aligned}
$$

由于 $\hat{X}=a$， 故 $E(\mu-\hat{X})^2=(\mu-\hat{X})^2$， $E(X-\mu)(\mu-\hat{X})=(\mu-\hat{X})E(X-\mu)=0$，再注意到 $E(X-\mu)^2=\sigma^2$，我们有

$$E(X-\hat{X})^2=\sigma^2+(\mu-\hat{X})^2.$$

由此立知当 $\hat{X}=\mu$ 时均方误差最小，即在关于 X 没有进一步信息可用时，对随机变量 X 的最小均方误差预报是它的均值 $\mu=EX$。

例 4.20 现设 X 和 Y 为有有限方差和已知联合分布的两个随机变量。假定根据 Y 的观察值来预报 X。这种例子在实际生产中经常见到，如根据已知的钢样品的硬度来预报它的强度，根据有随机误差的测量值如压力，温度等物理量来预报它的真值等等。

我们假定 Y 是可观察的，而对 X 没有进一步的信息（指没有比联合分布更多的其他的信息）。我们允许有有限方差的任一 Y 的函数 $\hat{X}=f(Y)$ 作为 X 的预报，即预报空间 $\mathscr{H}=\{\hat{X}=f(Y);f$ 为任一函数且 $\mathrm{Var}(\hat{X})<\infty\}$。

在观察到 $Y=y$ 后，我们可以得到最小均方误差预报是在 $Y=y$ 下 X 的条件均值 $\mu_{X|Y}=E(X|Y)$。这是因为

$$
\begin{aligned}
E(X-\hat{X})^2 = E(X-\mu_{X|Y})^2 &+ 2E(X-\mu_{X|Y})(\mu_{X|Y}-\hat{X}) \\
&+ E(\mu_{X|Y}-\hat{X})^2,
\end{aligned}
$$

在上式右边第二项中用条件期望平滑公式，并注意到

$$E[(X-\mu_{X|Y})|Y]=E(X|Y)-\mu_{X|Y}$$

$$= \mu_{X|Y} - \mu_{X|Y} = 0,$$

故

$$\begin{aligned}
&E(X-\mu_{X|Y})(\mu_{X|Y}-\hat{X})\\
&= E_Y\{E[(X-\mu_{X|Y})(\mu_{X|Y}-\hat{X})|Y]\}\\
&= E_Y\{(\mu_{X|Y}-\hat{X})E[(X-\mu_{X|Y})|Y]\}\\
&= 0,
\end{aligned}$$

由此得

$$E(X-\hat{X})^2 = E(X-\mu_{X|Y})^2 + E(\hat{X}-\mu_{X|Y})^2,$$

当 $\hat{X} = E(X|Y)$时上式取最小值.

例 4.21 $\hat{X} = E(X|Y)$的计算需要 X 和 Y 联合分布的全部知识。即使可以利用这一知识，这一公式也往往由于太繁杂而得不出实际结果. 此外，在平稳过程的研究中我们仅假定过程有二阶矩，而没有进一步关于联合分布的知识，因此要求有一种预报理论，在仅知道前二阶矩的情况下也能导出简单的预报公式，方法之一是线性预报.设 X,Y 分别有均值 μ_X,μ_Y 和方差 σ_X^2,σ_Y^2，它们的协方差为 $\sigma_{X,Y} = E(X-\mu_X)(Y-\mu_Y)$。我们限制 X 的预报只能是 Y 的线性函数,即 $\hat{X} = a + bY$或等价地 $\hat{X} = a + b(Y-\mu_Y)$这种形式,这儿 a 和 b 为任意实数. 由于允许的预报类比 例 4.20小，故最终算得的最小均方预报误差不可能比例4.20中算得的更小，完全可能要大一点。但由此得到的线性预报公式是简单的。我们下面证明最佳系数可以不用联合分布知识，而只用给定的矩知识来确定,且 X 根据 Y 的最佳线性预报为

$$\hat{X}^* = \mu_X + \frac{\sigma_{XY}}{\sigma_Y^2}(Y-\mu_Y).$$

下面来证明这一公式。 令 $\hat{X}$ 为任一线性预报. 即 $\hat{X} \in \mathscr{H} = \{a + b(Y-\mu_Y), a, b \text{ 为实数}\}$. 注意到 $\hat{X}^* - \hat{X}$ 仍是 X 的线性预报，故 $\hat{X}^* - \hat{X} \in \mathscr{H}$，且

$$\hat{X}^* - \hat{X} = a' + b'(Y-\mu_Y),$$

其中 $a' = \mu_X - a$，$b' = (\sigma_{XY}/\sigma_Y^2) - b$. 同前，我们计算

$$E(X-\hat{X})^2=E(X-\hat{X}^*)^2+2E(X-\hat{X}^*)(\hat{X}^*-\hat{X})$$
$$+E(\hat{X}^*-\hat{X})^2.$$

我们仍然可算得上式右边的交叉项为零。确实，

$$\begin{aligned}
&E(X-\hat{X}^*)(\hat{X}^*-\hat{X})\\
&=E\left[\left\{(X-\mu_X)-\frac{\sigma_{XY}}{\sigma_Y^2}(Y-\mu_Y)\right\}\left\{a'+b'(Y-\mu_Y)\right\}\right]\\
&=a'E\left[(X-\mu_X)-\frac{\sigma_{XY}}{\sigma_Y^2}(Y-\mu_Y)\right]\\
&\quad+b'E\left[(X-\mu_X)(Y-\mu_Y)-\frac{\sigma_{XY}}{\sigma_Y^2}(Y-\mu_Y)^2\right]\\
&=0+b'\left(\sigma_{XY}-\frac{\sigma_{XY}}{\sigma_Y^2}\sigma_Y^2\right)=0,
\end{aligned}$$

因此

$$E(X-\hat{X})^2=E(X-\hat{X}^*)^2+E(\hat{X}^*-\hat{X})^2.$$

当 $\hat{X}=\hat{X}^*$ 时，上式在一切线性预报中取最小值。故 $\hat{X}^*$ 是所述的线性最小均方误差预报。

有一种重要的特殊情况，此时由例 4.21 算得的预报就等于由条件期望公式在例 4.20 中算得的预报，这就是 X 和 Y 有联合正态分布。由 1.2 节和正态性可以算得，在给定 $Y=y$ 下 X 的条件期望是

$$E(X|Y=y)=\mu_X+\frac{\sigma_{XY}}{\sigma_Y^2}(y-\mu_Y).$$

注意到这个最佳的预报关于 y 是线性的。因此如果 X 和 Y 有联合正态分布，则给定 Y 时 X 的最小均方误差预报事实上就是线性最小均方误差预报。

在这些例子中，为证明给定的预报 $\hat{X}^*$ 是最优的，关键是证明交叉乘积项的期望 $E(X-\hat{X}^*)(\hat{X}^*-\hat{X})$ 为零。我们可以把上述推理总结为关于最小均方误差预报的预报定理：

定理 4.5 设 X 为有有限二阶矩的随机变量。$\mathscr{H}$ 为允许的

预报 $\hat{X}$ 所组成的空间，即$\mathscr{H}=\{\hat{X}: E\hat{X}^2<\infty\}$。假设$\mathscr{H}$是如下意义下的线性空间。若 $\hat{X}_1$ 和 $\hat{X}_2$ 都是 X 的预报，a 为任一实数，则 $a\hat{X}_1+\hat{X}_2$ 仍是 X 的一个允许预报，即 $a\hat{X}_1+\hat{X}_2\in\mathscr{H}$。则预报 $\hat{X}^*$ 有最小均方误差当且仅当对每个预报 $U\in\mathscr{H}$，有

$$E[(X-\hat{X}^*)U]=0。\tag{4.42}$$

此时我们称 $\hat{X}^*$ 为最小均方误差预报，简称最佳预报。

证明见附录 A。

定理 4.5 有个直观的几何解释。我们知道在有限维欧氏空间中，可以定义内积。两个向量相互垂直当且仅当它们的内积为零。一个向量 x 与一个超平面垂直当且仅当 x 与超平面中任一向量垂直。现设 $\mathscr{H}$ 由所有二阶矩有限的随机变量所组成。我们把 $E(XY)$ 定义为两个随机变量 X 和 Y 的内积。如果 $E(XY)=0$，则称随机变量 X 和 Y "垂直"，换句话说，若两个零均值的随机变量不相关，则对应的几何语言是"垂直"。类似于欧氏空间，若随机变量与一族随机变量的任一线性组合都垂直，则称 X 与这族随机变量所生成的随机变量空间垂直。在这个意义上，定理 4.5 就是说如果 $\hat{X}^*$ 是 X 在预报空间 $\mathscr{H}$ 上的最小均方误差预报，则$X-\hat{X}^*$ 与空间 $\mathscr{H}$ 垂直。但 $\hat{X}^*$ 在 $\mathscr{H}$ 中，所以上述说法等价于 $\hat{X}^*$ 是 X 在空间 $\mathscr{H}$ 上的投影。因此定理 4.5 又称为投影定理。图 4.6 将帮助导出几何直观。

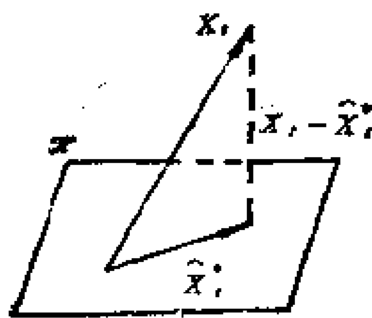


图 4.6 证明在子空间 $\mathscr{H}$中向量 X^* 最接近 X 的主要条件是 $X-X^*$ 与 $\mathscr{H}$ 中所有向量垂直

在最小均方误差预报中，如果把预报类仅限于 $\mathscr{H}$ 中随机变量的线性组合，则所得到的预报称为线性最小均方误差预报，简称为线性最佳预报。一般记为 $\hat{E}(X|\mathscr{H})$。如果$\mathscr{H}$ 仅由有限个随机变量 $X_1,\cdots,X_n$生成，即$\mathscr{H}=\{a_1X_1+\cdots+a_nX_n;\ a_1,\cdots,\ a_n$ 为任意实数$\}$。记随机向量 $Y=(X_1,\ \cdots,\ X_n)$，Y' 为Y的转置，则预报 $\hat{E}(X|\mathscr{H})$ 可写为

$\hat{E}(X|Y)$。对照有限维欧氏空间中向量投影的表达式，我们可以得到线性最佳预报的明显表达形式：

$$\hat{E}(X|\boldsymbol{Y}) = EX + \mathrm{Cov}(X,\boldsymbol{Y})(\mathrm{Var}(\boldsymbol{Y}))^{-1}(\boldsymbol{Y}' - E\boldsymbol{Y}'). \tag{4.43}$$

（推导见附录 B）。

现在我们用定理 4.5 的准则再次推出例4.21中的最佳预报。由于 X 仅为 Y 的线性函数，故 $\boldsymbol{Y}=Y$。由 4.43 知

$$\begin{aligned}\hat{E}(X|\boldsymbol{Y}) &= EX + \mathrm{Cov}(X,Y)(\mathrm{Var}Y)^{-1}(Y - EY)\\ &= \mu_X + \frac{\sigma_{XY}}{\sigma_Y^2}(Y - EY).\end{aligned}$$

4.4.2 平稳序列的预报

设$\{X_n,\ n=\cdots,-1,0,1\}$为一平稳序列。在时刻 n 我们要预报X_{n+l}的值，这称为 l 步预报。在实际中，到时刻 n 为止只能获得有限个数据，不妨设为 X_1，…，X_n。我们考虑 X_{n+l} 的线性最佳预报（这意味着这种预报只能是 X_1，…，X_n 的线性组合）。为说明它由数据 X_1，…，X_n 得到，记为 $\hat{X}_{n+l|n,1}$。由（4.43），得

$$\hat{X}_{n+l|n,1} = EX_{n+l} + \mathrm{Cov}(X_{n+l},\boldsymbol{Y}_{n,1}) \cdot (\mathrm{Var}(\boldsymbol{Y}_{n,1}))^{-1}(\boldsymbol{Y}'_{n,1} - E\boldsymbol{Y}'_{n,1}), \tag{4.44}$$

其中 $\boldsymbol{Y}_{n,1}=(X_1,\cdots,X_n)$。假如我们得到一个新数据，即$X_{n+1}$的观察值。并用 $X_1,\cdots,X_n,X_{n+1}$ 来作线性最佳预报，记为$\hat{X}_{n+l+1|n+1,1}$。显然，当 n 增大时，预报公式(4.44)不断变化。除非能找到递推公式并使用计算机，否则是很难作预报的。

由于 $\hat{X}_{n+l|n,1}$ 随 n 的变化不构成 n 的平稳序列，所以称之为非平稳预报。这种预报只有当数据量不大时才有使用价值，因此这儿暂不讨论。

为了克服上述非平稳预报的缺点，我们假设在时刻 n 知道了 X_n，X_{n-1}，…直至 $-\infty$ 的全部数据，并用这无穷个数据来预报

X_{n+l}。这样得到的预报形式固定，也便于使用。把这种预报记为$\hat{X}_{n+l\,n,-\infty}$，简记为$\hat{X}_{n+l\,n}$。当然，在实际中，我们不可能得到从$-\infty$到n的全部数据，那末如何来实现预报呢？在以后的预报公式中可以看到，即使数据从X_1开始，只要做预报的时刻n适当大，则在预报公式中，零时刻以前的数据实际上对预报值已不起什么作用。因而可以较随意地取值，比如说都可以取作0。这样就归结为知道从$-\infty$到n全部数据这种情况。下面介绍最常见的自回归模型和滑动平均模型的线性最佳预报。

1．自回归模型的线性最佳预报

从现在开始，我们总假定$\{\varepsilon_n, n=\cdots,-1,0,1,\cdots\}$是均值为0，方差为$\sigma^2$的白噪声序列。$\mathscr{H}_n$表示由$\{\varepsilon_k,\ k\leqslant n\}$所生成的线性空间，即$\mathscr{H}_n$中任一元素可表为诸$\varepsilon_k$，$k\leqslant n$的线性组合。

设随机序列$\{X_n, n=\cdots,-1,0,1\cdots\}$满足

$$X_n=\alpha X_{n-1}+\varepsilon_n,\quad n=\cdots,-1,0,1,\cdots \tag{4.45}$$

由习题17知

$$X_n=\sum_{k=0}^{\infty}\alpha^k\varepsilon_{n-k}, \tag{4.46}$$

$$EX_n=0,\quad \sigma_X^2=EX_n^2=\frac{\sigma^2}{1-\alpha^2},$$

$$\mathrm{Cov}(X_n,\ X_{n+h})=\alpha^h\sigma_X^2. \tag{4.47}$$

从而$\{X_n,\ n=\cdots,-1,0,1,\cdots\}$为平稳序列，称为一阶自回归模型(Autoregressive Model)，记为$AR(1)$。关于X_n的表达式还有另一种算子推导法。设z表示一种线性后移算子。它对任一序列$\{\eta_n, n=\cdots,-1,0,1,\cdots\}$作用如下：

$$z\eta_n=\eta_{n-1},$$

$$z^2\eta_n=z(z\eta_n)=z\eta_{n-1}=\eta_{n-2}.$$

从而对一切$k=1,2,\cdots$，有

$$z^k\eta_n=z^{k-1}(z\eta_n)=z^{k-1}\eta_{n-1}=\eta_{n-k}.$$

此外，还成立

$$(az^k+bz^k)\eta_n=az^k\eta_n+bz^k\eta_n=a\eta_{n-k}+b\eta_{n-k}.$$

若记 $z^0=1$ 为恒等算子，则(4.45)可写为

$$X_n-\alpha zX_n=\varepsilon_n$$

或 $$(1-\alpha z)X_n=\varepsilon_n.$$

形式上可得到

$$X_n=(1-\alpha z)^{-1}\varepsilon_n=\sum_{k=0}^{\infty}\alpha^k z^k\varepsilon_n=\sum_{k=0}^{\infty}\alpha^k\varepsilon_{n-k},\tag{4.48}$$

即同样得到 X_n 的表达式(4.45)。可以证明只要(4.48)最右边级数收敛(指均方收敛)。则表达式 (4.48) 是合理的。由于 ε_n 为白噪声，故 $E\left(\sum_{k=0}^{\infty}\alpha^k\varepsilon_{n-k}\right)^2=\sum_{k=0}^{\infty}\alpha^{2k}\sigma^2$，于是只要 $|\alpha|<1$ 就能保证级数 $\sum_{k=0}^{\infty}\alpha^k$ 收敛。$AR(1)$的谱密度函数为

$$S(\omega)=\frac{1-\rho^2}{1-2\rho\cos\omega+\rho^2}.$$

现讨论 $AR(1)$的 l 步线性最佳预报。由(4.46)，

$$\begin{aligned}X_{n+l}&=\sum_{k=0}^{\infty}\alpha^k\varepsilon_{n+l-k}\\&=(\varepsilon_{n+l}+\alpha\varepsilon_{n+l-1}+\cdots+\alpha^{l-1}\varepsilon_{n+1})+\alpha^l\sum_{k=0}^{\infty}\alpha^k\varepsilon_{n-k}.\end{aligned}\tag{4.49}$$

由于$\{\varepsilon_n,n=\cdots,-1,0,1,\cdots\}$ 为白噪声。所以 $\varepsilon_{n+l},\cdots,\varepsilon_{n+1}$ 与$\{\varepsilon_k,k\leqslant n\}$正交，从而与 $\mathscr{H}_n$ 正交。但 ε_n 是 X_n 和 X_{n-1} 的线性组合，所以由 $\{X_k,\ k\leqslant n\}$ 生成的预报空间与 $\mathscr{H}_n$ 相同。因此 X_{n+l} 可以分为两部分。第一部分$(\varepsilon_{n+l}+\alpha\varepsilon_{n+l-1}+\cdots+\alpha^{l-1}\varepsilon_{n+1})$与 $\mathscr{H}_n$ 正交，第二部分 $\alpha^l\sum_{k=0}^{\infty}\alpha^k\varepsilon_{n-k}$ 是 $\mathscr{H}_n$ 中的一个元素。这种分解称为 Wold 分解。因此由定理 4.5 知 X_{n+l} 的线性最佳预报应是 X_{n+l} 在 $\mathscr{H}_n$ 上的投影。即

$$\hat{X}_{n+l|n}=\alpha^l\sum_{k=0}^{\infty}\alpha^k\varepsilon_{n-k}=\alpha^lX_n.$$

由(4.46)其预报误差为

$$X_{n+l}-\hat{X}_{n+l|n}=\varepsilon_{n+l}+\alpha\varepsilon_{n+l-1}+\cdots+\alpha^{l-1}\varepsilon_{n+1},\tag{4.50}$$

$$E(X_{n+l}-\hat{X}_{n+l|n})^2=\sum_{k=0}^{l-1}\alpha^{2k}\sigma^2=\frac{1-\alpha^{2l}}{1-\alpha^2}\sigma^2=(1-\alpha^{2l})\sigma_X^2.$$

由上式知，当 l 充分大时．$AR(1)$ 的 l 步预报误差与自回归序列本身的方差差不多。

例 4.22 在某个计算机的固定程序中，包括如下简单迭代计算

$$y_n=\alpha y_{n-1},\qquad |\alpha|<1 \tag{4.51}$$

由于计算机的字长有限，每次计算（4.51）式时都会有舍入误差。若以 y_n^0 和 y_n 分别表示计算机的计算值和真实值，那么差 $x_n=y_n^0-y_n$ 就是一个误差序列，我们来分析它的变化规律．设在计算 y_n^0 时产生的舍入误差为 ε_n，则计算值 y_n^0 为

$$y_n^0=\alpha y_{n-1}^0+\varepsilon_n, \tag{4.52}$$

其中 αy_{n-1}^0 表示 α 与 y_{n-1}^0 的真实乘积值，由于机器有舍入误差 ε_n，所以 y_n^0 有(4.52)的形式。联合(4.51)和(4.52)两式就有

$$x_n=\alpha x_{n-1}+\varepsilon_n,$$

其中舍入误差近似为$(-0.5,0.5)$之间的均匀分布，且 $\{\varepsilon_n\}$ 两两不相关。所以计算误差序列满足 $AR(1)$模型。试对 $\alpha=0.8$ 作误差的 l 步预报并求预报方差。

解 由于 ε_n 服从$(-0.5,0.5)$上的均匀分布，故

$$\sigma^2=1/12,\quad \sigma_X^2=\sigma^2/(1-\alpha^2)=25/108.$$

由此知

$$\hat{X}_{n+l|n}=\alpha^l X_n=0.8^l X_n,$$

$$E(X_{n+l}-\hat{X}_{n+l|n})^2=25(1-0.8^{2l})/108.$$

如果设 $X_0=10^{-1}$，则下步误差预报为 0.08，均方差为 0.290 若作二步预报，则 $\hat{X}_{2|0}=0.064$，均方差为 0.370。

模型(4.45)可以进一步推广如下，若随机序列 $\{X_n,n=\cdots,-1,0,1,\cdots\}$ 满足

$$X_n=\alpha_1X_{n-1}+\alpha_2X_{n-2}+\cdots+\alpha_pX_{n-p}+\varepsilon_n,$$
$$n=\cdots,-1,0,1,\cdots. \tag{4.53}$$

则$\{X_n\}$称为p阶自回归模型，记为$AR(p)$。如前所述，若以z记后移算子，则(4.53)可改写为

$$(1-\alpha_1 z-\alpha_2 z^2-\cdots-\alpha_p z^p)X_n=\varepsilon_n. \tag{4.54}$$

形式上，(4.54)可以写为

$$X_n=(1-\alpha_1 z-\cdots-\alpha_p z^p)^{-1}\varepsilon_n.$$

如果多项式$1-\alpha_1 z-\cdots-\alpha_p z^p$的根都在单位圆外，则$(1-\alpha_1 z-\cdots-\alpha_p z^p)^{-1}$可以展成$z$的收敛幂级数($|z|<1$)，记为$\psi(z)=\sum\limits_{k=0}^{\infty}\beta_k z^k$，$\beta_0=1$。则同样可以从数学上严格证明(4.54)可以写成如下形式：

$$X_n=\psi(z)\varepsilon_n=\sum_{k=0}^{\infty}\beta_k z^k\varepsilon_n=\sum_{k=0}^{\infty}\beta_k\varepsilon_{n-k}. \tag{4.55}$$

由此不难得到

$$EX_n=0,\quad EX_n^2=\sum_{k=0}^{\infty}\beta_k^2\sigma^2,$$

$$\mathrm{Cov}(X_n,X_{n+h})=\sigma^2\sum_{k=0}^{\infty}\beta_k\beta_{k+h}. \tag{4.56}$$

从而$AR(p)$也是平稳序列。现考虑其线性最佳预报。同$p=1$的场合，由$\{X_k,k\leqslant n\}$所构成的线性空间与由$\{\varepsilon_k,k\leqslant n\}$生成的线性空间$\mathscr{H}_n$相同。而$X_{n+1}=(\alpha_1 X_n+\cdots+\alpha_p X_{n+1-p})+\varepsilon_{n+1}$，其中$\varepsilon_{n+1}$与$\mathscr{H}_n$垂直，而$\alpha_1 X_n+\cdots+\alpha_p X_{n+1-p}$是$\mathscr{H}_n$中的元素，因此由定理4.5得一步预报为

$$\hat{X}_{n+1|n}=\alpha_1 X_n+\cdots+\alpha_p X_{n+1-p}. \tag{4.57}$$

其预报误差为$X_{n+1}-\hat{X}_{n+1|n}=\varepsilon_{n+1}$，故

$$E(X_{n+1}-\hat{X}_{n+1|n})^2=E\varepsilon_{n+1}^2=\sigma^2.$$

关于多步预报，由定理4.5也可以推出

$$\hat{X}_{n+l|n}=\alpha_1\hat{X}_{n+l-1|n}+\alpha_2\hat{X}_{n+l-2|n}+\cdots+\alpha_p\hat{X}_{n+l-p|n}. \tag{4.58}$$

在这儿注意如下事实：若$m\leqslant p$，则$\hat{X}_{n+m-p|n}$就是X_{n+m-p}。这是因为X_{n+m-p}在$\mathscr{H}_n$中，故它的投影即为本身。

从(4.58)式可以看出，对$AR(p)$模型，在时刻n无论作多

少步预报，仅用到 $X_n, X_{n-1}, \cdots, X_{n-p+1}$ 这 p 个数据，再早的数据对预报已不起作用.

关于预报误差由(4.55)，

$$X_{n+l}=\sum_{k=0}^{\infty}\beta_k\varepsilon_{n+l-k}.$$

因此 $\hat{X}_{n+l|n}$ 作为 X_{n+l} 在 $\mathscr{H}_n$ 上的投影．应有

$$\hat{X}_{n+l|n}=\sum_{k=0}^{\infty}\beta_{l+k}\varepsilon_{n-k}. \tag{4.59}$$

此式与(4.58)并不矛盾，只不过表达形式不同而已．由 (4.59) 及(4.55)得

$$X_{n+l}-\hat{X}_{n+l|n}=\sum_{k=0}^{l-1}\beta_k\varepsilon_{n+l-k},$$

$$E(X_{n+l}-\hat{X}_{n+l|n})=\sum_{k=0}^{l-1}\beta_k^2\sigma^2. \tag{4.60}$$

例 4.23 考虑如图带反馈的线性滤波器．此滤波器表示先把放大 $-\alpha$ 倍的前一时刻的输出和放大 $-\beta$ 倍的前二时刻的输出叠加，然后再和 t 时刻的脉冲输入 ε_t 叠加以得出在 t 时刻的输出．即

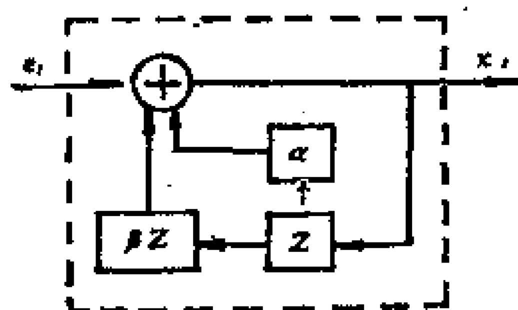


图4.7

$$X_t=\varepsilon_t+(-\alpha X_{t-1}-\beta X_{t-2})$$
$$=\varepsilon_t-\alpha X_{t-1}-\beta X_{t-2}.$$

由于输出反馈到输入端，因此在每一时刻的输入对今后的影响将持续下去．这是一个 $AR(2)$ 模型．现设 $\alpha=-1$，$\beta=\frac{1}{2}$，$\{\varepsilon_n\}$是均值为零，方差为 1 的白噪声序列．求在时刻 n 时的协方差函数及 l 步预报．

解 由于 $1-z+\frac{1}{2}z^2$ 的根为 $\sqrt{2}\,e^{\pm\frac{\pi}{4}j}$，在单位圆外，故若令 $b=\frac{1}{\sqrt{2}}e^{j\frac{\pi}{4}}$，则

$$\left(1-z+\frac{1}{2}z^2\right)^{-1}=\frac{1}{j}\left(\frac{1}{1-bz}-\frac{1}{1-\bar{b}z}\right)$$

$$=\frac{1}{j}\sum_{k=0}^{\infty}(b^k-\bar{b}^k)z^k=\sum_{k=0}^{\infty}2^{1-\frac{k}{2}}\left(\frac{e^{\frac{\pi k}{4}j}-e^{-\frac{\pi k}{4}j}}{2j}\right)z^k$$

$$=\sum_{k=0}^{\infty}2^{1-\frac{k}{2}}\sin\left(\frac{\pi k}{4}\right)z^k\triangleq\sum_{k=0}^{\infty}\beta_k z^k.$$

故由(4.56)

$$R(h)=\sum_{k=0}^{\infty}\beta_k\beta_{k+h}=\sum_{k=0}^{\infty}2^{1-\frac{k}{2}}2^{1-\frac{k+h}{2}}\sin\frac{\pi k}{4}\sin\frac{\pi(k+h)}{4}$$

$$=\frac{1}{3}\cdot 2^{4-h/2}\left(\sin\frac{\pi h}{2}+\frac{1}{2}\cos\frac{\pi}{2}h\right).$$

由(4.58)式，我们得

$$\hat{X}_{n+l|n}=-\hat{X}_{n+l-1|n}+\frac{1}{2}\hat{X}_{n+l-2|n},\qquad l\geqslant-1.\qquad(4.60)$$

这是一个差分方程。设 b，$\bar{b}$ 是$1-z+\frac{1}{2}z^2$的两根，则 $\hat{X}_{n+l|n}$ 可表为

$$\hat{X}_{n+l|n}=c_1b^{-l}+c_2\bar{b}^{-l},\quad l\geqslant-1.$$

其中常数 c_1 和 c_2 由下列方程组确定：

$$c_1+c_2=X_n,$$
$$c_1b+c_2\bar{b}=X_{n-1}.$$

解此方程组得 $c_1=\frac{1-j}{2}X_n+\frac{j}{2}X_{n-1}$，$c_2=\frac{1+j}{2}X_n-\frac{j}{2}X_{n-1}$.

故

$$\hat{X}_{n+l|n}=2^{-\frac{l-1}{2}}X_n\cos\frac{l-1}{4}\pi-2^{-\frac{l}{2}}X_{n-1}\sin\frac{l}{4}\pi,\quad l\geqslant-1.$$

可以看出，预报函数以指数速度趋于0。

例 4.24 在关于太阳黑子年平均数，全国铁路客流量，跟踪雷达误差特性的鉴定问题以及惯性陀螺的随机漂移等方向的研究中，常用 $AR(2)$模型来描述。今以跟踪雷达误差特性的鉴定问

题为例。一部目标跟踪雷达质量的好坏，最终要通过它跟踪目标时的测量误差来衡量。为了鉴定这种雷达，首先用它跟踪飞行目标(如导弹)，记录其飞行轨道数据，即目标和雷达的瞬时径向距离 $D(t)$，仰角 $\theta(t)$ 和方位角 $\varphi(t)$。现等间隔采样，得到三列数据 $\{D_k,\theta_k,\varphi_k,\ k=1,2,\cdots\}$，其中 $D_k=D(kh)$，$\theta_k=\theta(kh)$，$\varphi_k=\varphi(kh)$，h 为采样间隔。与此同时，用更精确的观测设备(例如经纬仪照象设备)观测同一目标，并记录以上三个相应的坐标序列。记为 $D_k^0,\theta_k^0,\varphi_k^0,k=1,2,\cdots$，这些量相对于 D_k，θ_k 和 φ_k 而言，可以近似看作目标的真实轨道坐标数据。于是 $\Delta D_k=D_k-D_k^0$，$\Delta\theta_k=\theta_k-\theta_k^0$，$\Delta\varphi_k=\varphi_k-\varphi_k^0$ 就可以看作雷达对此目标跟踪时的测量误差序列。根据大量的数据分析结果表明，在雷达跟踪同一空域的飞行目标时，其跟踪误差序列 $\Delta D_k,\Delta\theta_k,\Delta\varphi_k$，$k=1,2,\cdots$ 都可以近似地用 $AR(2)$ 模型来描述。只不过每个模型对应的参数有所不同罢了。

最后，我们还应注意到，如果 $\{\varepsilon_n,n=\cdots,-1,0,1\cdots\}$ 是高斯白噪声，则预报 $\hat{X}_{n+l|n}$ 是一个正态随机变量，其预报误差 $X_{n+l}-\hat{X}_{n+l|n}$ 也是一个正态随机变量，这在作进一步的分析中是有用的。

2．滑动平均模型的预报

若随机序列 $\{X_n,n=\cdots,-1,0,1,\cdots\}$ 满足

$$X_n=\varepsilon_n-\beta_1\varepsilon_{n-1}-\cdots-\beta_q\varepsilon_{n-q},\quad n=\cdots,-1,0,1,\cdots. \qquad (4.61)$$

则 $\{X_n\}$ 称为 q 阶滑动平均模型，记为 $MA(q)$（MA 是 moving average 的缩写）。

第一节中例 4.7 证明了滑动平均是平稳序列。

例 4.25 考虑如图 4.8 所示的线性滤波器。脉冲输入 ε_t 和前一时刻脉冲输入 ε_{t-1} 被放大 $-\beta$ 倍的量叠加成 X_t 输出。

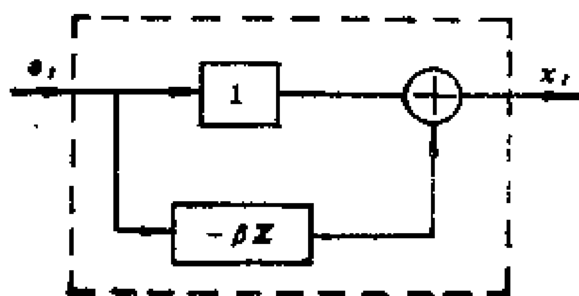


图 4.8

图中 Z 表示信号的一步延迟环节，即信号在时刻 t 的输出是 $t-1$ 时输入的信号。用式子表示此滤波器。则有

$$X_t=\varepsilon_t-\beta\varepsilon_{t-1},$$

这儿 β 可正可负。这是一个 $MA(1)$模型。

为了讨论 $MA(q)$ 的预报公式，我们仍用前已用过的后移算子记号 z。(4.61)可改写为

$$X_n=(1-\beta_1 z-\cdots-\beta_q z^q)\varepsilon_n. \tag{4.62}$$

如果 $l>q$，由上式及定理 4.5 易知 $\hat{X}_{n+l|n}=0$。当 $l\leqslant q$ 时，由于 $\{\varepsilon_n\}$为白噪声，故我们有

$$\hat{X}_{n+l|n}=-\beta_l\varepsilon_n-\cdots-\beta_q\varepsilon_{n-q}. \tag{4.63}$$

但由于$\{\varepsilon_n\}$不是数据，因此不便直接使用上式预报。但若 $1-\beta_1 z-\cdots-\beta_q z^q$ 的根都在单位圆外，利用后移算子，(4.62)可改写为

$$\begin{aligned}\varepsilon_n&=(1-\beta_1 z-\cdots-\beta_q z^q)^{-1}X_n\\&=\sum_{k=0}^{\infty}\alpha_k z^k X_n=\sum_{k=0}^{\infty}\alpha_k X_{n-k}.\end{aligned} \tag{4.64}$$

其中$\sum_{k=0}^{\infty}\alpha_k z^k$ 是$(1-\beta_1 z-\cdots-\beta_q)^{-1}$ 的幂级数展开，$\alpha_0=1$。(4.64)也称为无穷阶自回归模型。由此知滑动平均模型可以写成无穷阶自回归的形成。首先讨论一步预报。由(4.64)，我们有

$$X_{n+1}=-\sum_{k=1}^{\infty}\alpha_k X_{n-k+1}+\varepsilon_{n+1},$$

其中$-\sum_{k=1}^{\infty}\alpha_k X_{n-k+1}$ 为$\{X_k, k\leqslant n\}$的线性组合，所以是 $\mathscr{H}_n$ 中的元素，而 ε_{n+1} 与 $\mathscr{H}_n$ 正交。所以由定理 4.5 有

$$\hat{X}_{n+1|n}=-\sum_{k=1}^{\infty}\alpha_k X_{n-k+1}, \tag{4.65}$$

$$E(X_{n+1}-\hat{X}_{n+1|n})^2=E\varepsilon_{n+1}^2=\sigma^2.$$

关于多步预报，设 $l\geqslant 2$，由预报定理，$\hat{X}_{n+l|n}$ 必可表为$\{X_k, k\leqslant n\}$的线性组合，即

$$\hat{X}_{n+l|n}=-\sum_{k=1}^{\infty}\alpha_k^{(l)}X_{n+1-k}. \tag{4.66}$$

另一方面，由(4.64)，

$$X_{n+l}=-\sum_{k=1}^{l-1}\alpha_k X_{n+l-k}-\sum_{k=l}^{\infty}\alpha_k X_{n+l-k}+\varepsilon_{n+l}. \quad (4.67)$$

由定理 4.5，$\hat{X}_{n+l|n}$ 是 X_{n+l} 在 $\mathscr{H}_n$ 上的投影。而在(4.67)中，$-\sum\limits_{k=l}^{\infty}\alpha_k X_{n+l-k}$ 是 $\mathscr{H}_n$ 中的元素，因此 X_{n+l} 在 $\mathscr{H}_n$ 上的投影为

$$\hat{X}_{n+l|n}=-\sum_{k=1}^{l-1}\alpha_k\hat{X}_{n+l-k|n}-\sum_{k=l}^{\infty}\alpha_k X_{n+l-k}. \quad (4.68)$$

上式可以作为递推公式，对 $l=1,2,\cdots$ 递推得出预报 $\hat{X}_{n+1|n}$，$\hat{X}_{n+2|n}$，…。把(4.66)代入(4.68)(代入时先把 l 换成 $l-1$，$l-2,\cdots,1$)，然后比较两边 X_{n-j} 前的系数，得

$$\alpha_k^{(l)}=\alpha_{k+l-1}-\sum_{i=1}^{l-1}\alpha_i\alpha_k^{(l-i)},\ l>1. \quad (4.69)$$

利用此式，先取 $l=1$ 时 $\alpha_k^{(1)}$ 的初值(见(4.65)式)，递推得 $\alpha_k^{(l)}$，$l=2,3,\cdots$，再代入(4.66)，得 $\hat{X}_{n+l|n}$，$l=2,3,\cdots$。

上面的预报要求我们有无穷个数据，但实际中只能得到有限个，如何解决这一矛盾？由(4.64)，由 α_k 的定义知 $\sum\limits_{k=0}^{\infty}|\alpha_k|^2<\infty$，因而 $k\to\infty$ 时，$\alpha_k\to 0$，因此如果 $\{X_k\}$ 的取值在某个有界范围中，则只要 m 充分大，$-\sum\limits_{k=m+1}^{\infty}\alpha_k X_{n+1-k}$ 就很小，从而对 $\hat{X}_{n+1|n}$ 的影响也就不大。通常可取为 0，这一点已在本节开始时就已说明。

至于 l 步预报的误差方差，由(4.67)和(4.68)当 $l\leqslant q$ 时得

$$X_{n+l}-\hat{X}_{n+l|n}=\varepsilon_{n+l}-\beta_1\varepsilon_{n+l-1}-\cdots-\beta_{l-1}\varepsilon_{n+1},$$
$$E(X_{n+l}-\hat{X}_{n+l|n})^2=(1+\beta_1^2+\cdots+\beta_{l-1}^2)\sigma^2. \quad (4.70)$$

例 4.26 设 $X_n=\varepsilon_n-\varepsilon_{n-1}+0.21\varepsilon_{n-2}$，求 X_n 的线性最佳预报以及预报误差方差，其中 σ^2 假定已知。

解 令 $B(z)=1-z+0.21z^2=(1-0.3z)(1-0.7z)$，

$$B^{-1}(z)=\frac{1}{1-0.3z}\cdot\frac{1}{1-0.7z}=\frac{1}{4}\left(\frac{7}{1-0.7z}-\frac{3}{1-0.3z}\right)$$
$$=\sum_{k=0}^{\infty}2.5[(0.7)^{k+1}-(0.3)^{k+1}]z^k.$$

由(4.64)得

$$\alpha_k = 2.5[(0.7)^{k+1} - (0.3)^{k+1}], \qquad k \geqslant 1.$$

由(4.69)我们有

$$\alpha_k^{(2)} = \alpha_{k+1} - \alpha_1\alpha_k = 0.25(7 \cdot (0.3)^{k+1} - 3(0.7)^{k+1}), \; k \geqslant 1.$$

如果要求相对误差不超过 0.05，则大约取 9 项就可以了，即只要有 9 个数据，就可以开始预报。

$$\hat{X}_{n+1|n} = -\sum_{k=1}^{\infty} \alpha_k X_{n+1-k} \approx \sum_{k=1}^{5} 2.5[(0.3)^{k+1} - (0.7)^{k+1}] X_{n+1-k},$$

$$\hat{X}_{n+2|n} = -\sum_{k=1}^{\infty} \alpha_k^{(2)} X_{n+1-k} \approx 0.25 \sum_{k=1}^{9} [3(0.7)^{k+1} - 7(0.3)^{k+1}] X_{n+1-k}.$$

$$\hat{X}_{n+l|n} = 0, \; l > 2.$$

一步预报误差方差为 σ^2。二步预报误差由(4.70)知为

$$(1+1)\sigma^2 = 2\sigma^2.$$

自回归模型和滑动平均模型的进一步推广是所谓的混合自回归——滑动平均模型。其一般形式为

$$X_n - \alpha_1 X_{n-1} - \cdots - \alpha_p X_{n-p} = \varepsilon_n - \beta_1 \varepsilon_{n-1} - \cdots - \beta_q \varepsilon_{n-q}, \qquad (4.71)$$

$$n = \cdots, -1, 0, 1, \cdots.$$

记为 $ARMA(p,q)$。如用前面多次使用的后移算子，可以把(4.71)改写为

$$\varphi(z) X_n = \psi(z) \varepsilon_n, \qquad n = \cdots, -1, 0, 1, \cdots. \qquad (4.72)$$

其中

$$\varphi(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p,$$

$$\psi(z) = 1 - \beta_1 z - \cdots - \beta_q z^q.$$

如果 $\varphi(z)$ 的根都在单位圆外，则 $\varphi^{-1}(z)$ 可以展成收敛的幂级数，从而形式上(4.72)可以记为

$$X_n = \varphi^{-1}(z)\psi(z)\varepsilon_n = \sum_{k=0}^{\infty} \gamma_k z^k \varepsilon_n = \sum_{k=0}^{\infty} \gamma_k \varepsilon_{n-k}. \qquad (4.73)$$

理论上可以严格证明当 $\varphi(z)$ 的根全部在单位圆外时，(4.73)式确实成立，由此可得

$$EX_n = 0, \; EX_n^2 = \sum_{k=0}^{\infty} \gamma_k^2 \sigma^2,$$

$$\mathrm{Cov}(X_n, X_{n+h}) = \sum_{k=0}^{\infty} \gamma_k \gamma_{k+h}.$$

从而模型(4.71)是平稳序列。用定理 4.5 也可以求得 X_n 的线性最佳预报。

在预测理论中还常常遇到如下两个问题。一是如果我们的观测数据$\{X_n\}$由于某种原因而残缺不全，例如在 $n\in[n_1,n_2]$ 中，我们缺了第 k 项数据 X_k，$n_1<k<n_2$，如何根据现有的数据去补全？这称为内插。第二个问题是在实际中我们得到的观测值 X_n 一般是由信号 s_n 和噪声 ε_n 两部分叠加而成。设 $X_n=s_n+\varepsilon_n$，$n=\cdots,-1,0,1,\cdots$ 如何根据n时刻之前的观测数据序列$\{X_k, k\leqslant n\}$把噪声 ε_n 滤掉，求出有用的信号 s_n。这称为滤波。第三个问题是在控制论中如何根据现有的测量数据，把过程控制在规定的范围内。这三个问题和预报问题有非常密切的联系。限于本书篇幅，不能作进一步的讨论了。

这儿所讨论的预报问题是在时间域范围内进行的。众所周知，对平稳过程，既可在时间域范围内研究，也可在频率域范围内研究，这要看具体问题来确定。因此也自然会想到在频率域范围内对频谱的估计问题。这些问题统称谱分析，读者对这方面感兴趣的话，可以参阅有关时间序列的书籍。

习　题　四

以下如果没有指明变量 t 的取值范围，一般视为 $t\in R$。平稳过程是指宽平稳过程。

1. 设 $X(t)=\sin Ut$，这儿 U 为$(0,2\pi)$上的均匀分布。

(a) 若 $t=1,2,\cdots$，证明$\{X(t), t=1,2,\cdots\}$是宽平稳但不是严平稳过程。

(b) 设 $t\in[0,\infty)$，证明$\{X(t), t\geqslant 0\}$既不是严平稳也不是宽平稳过程。

2．设$\{X_n, n=0,1,2\cdots\}$是平稳序列．定义$\{X_n^{(i)}, n=1,2,\cdots\}$ $i=1,2,\cdots$为

$$X_n^{(1)}=X_n-X_{n-1},$$
$$X_n^{(2)}=X_n^{(1)}-X_{n-1}^{(1)},$$
$$\vdots$$

证明这些序列仍是平稳序列。

3．设$X_n=\sum_{k=1}^{N}\sigma_k\sqrt{2}\cos(a_k n-U_k)$，这儿$\sigma_k$和$a_k$为正常数，$k=1,\cdots,N$；$U_1,\cdots,U_N$是$(0,2\pi)$上独立均匀分布随机变量。证明$\{X_n, n=0,\pm1,\cdots\}$是平稳过程。

4．设$A_k, k=1,2,\cdots,n$是n个实随机变量；ω_k，$k=1,2,\cdots,n$是n个实数．试问A_k以及A_k之间应满足怎样的条件才能使

$$Z(t)=\sum_{k=1}^{n}A_k e^{j\omega_k t}$$

是一个复的平稳过程。

5．设$\{X_n, n=1,2,\cdots\}$是一列独立同分布随机变量序列。$P(X_n=1)=p, P(X_n=-1)=1-p, n=1,2,\cdots$令$S_0=0, S_n=(X_1+\cdots+X_n)/\sqrt{n}, n=1,2,\cdots$求随机序列$\{S_n, n=1,2,\cdots\}$的协方差函数和相关函数。$p$取何值时此序列为平稳序列？

6．设$\{X(t)\}$是一个平稳过程，对每个$t\in R$，$X'(t)$存在。证明对每个给定的t，$X(t)$与$X'(t)$不相关，其中$X'(t)=\frac{dX(t)}{dt}$.

7．设$\{X(t)\}$是高斯过程，均值为0，协方差函数$R(\tau)=4e^{-2|\tau|}$．令$Z(t)=X(t+1)$，$W(t)=X(t-1)$。

(i) 求$E(Z(t)W(t))$和$E(Z(t)+W(t))^2$；

(ii) 求$Z(t)$的密度函数$f_Z(z)$及$P(Z(t)<1)$；

(iii) 求$Z(t)$，$W(t)$的联合密度$f_{Z,W}(z,\omega)$。

8．设$\{X(t), t\in R\}$是一个严平稳过程，ε为只取有限个值的随机变量。证明$\{Y(t)=X(t-\varepsilon), t\in R\}$仍是一个严平稳过程。

提示：对 n 用全概率公式。

9. 设 $\{X(t), t\in R\}$ 是一个严平稳过程。构造随机过程 Y 如下：$Y(t)=1$，若 $X(t)>0$；-1，若 $X(t)\leqslant 0$。证明 $\{Y(t), t\in R\}$ 是一个平稳过程。如果进一步假定 $\{X(t), t\in R\}$ 是高斯平稳过程，证明 $R_Y(\tau)$ 为 $\dfrac{2}{\pi}\arcsin(R_X(\tau)/R_X(0))$。

提示：只要证明 $P(X(t+\tau)X(t)>0)=\dfrac{1}{2}+\dfrac{\alpha}{\pi}$，其中 $\sin\alpha=R_X(\tau)/R_X(0)$。

10. 设 $\{X(t)\}$ 是一个复值平稳过程。证明

$$E|X(t+\tau)-X(t)|^2=2Re(R(0)-R(\tau))。$$

11. 设 $\{X(t)\}$ 是平稳高斯过程，协方差函数为 $R(\tau)$。证明

$$P(X'(t)\leqslant a)=\Phi\left(\frac{+a}{\sqrt{-R''(0)}}\right),$$

其中 $\Phi(\cdot)$ 为标准正态分布函数。

12. 设 $\{X(t)\}$ 为连续宽平稳过程，均值 m 未知，协方差函数为 $R(\tau)=ae^{-b|\tau|}$，$\tau\in R, a>0, b>0$。对固定的 $T>0$，令 $\overline{X}=T^{-1}\int_0^T X(s)\mathrm{d}s$。证明 $E\overline{X}=m$（即 $\overline{X}$ 是 m 的无偏估计）以及

$$\mathrm{Var}(\overline{X})=2a[(bT)^{-1}-(bT)^{-2}(1-e^{-bT})]。$$

提示：在上述条件下，期望号与积分号可以交换。

13. 设 $\{X(t)\}$ 为平稳过程，设 $X(t)$ 的 n 阶导数 $X^{(n)}(t)$ 存在。证明 $\{X^{(n)}(t)\}$ 是平稳过程。

提示：利用协方差函数性质 4。

14. 证明定理 4.1 中关于平稳序列均值的遍历性定理。提示：用Schwarz 不等式。

15. 如果 (X_1, X_2, X_3, X_4) 是均值为 0 的联合正态随机向量，则 $EX_1X_2X_3X_4=\mathrm{Cov}(X_1, X_2)\mathrm{Cov}(X_3, X_4)+\mathrm{Cov}(X_1, X_3)\mathrm{Cov}(X_2, X_4)+\mathrm{Cov}(X_1, X_4)\mathrm{Cov}(X_2, X_3)$。利用这个事实证

定理 4.3。

16. 设 X_0 为随机变量，其概率密度函数为

$$f(x)=\begin{cases}2x, & 0\leqslant x\leqslant 1,\\ 0, & \text{其他。}\end{cases}$$

设 X_{n+1} 在给定 $X_0,X_1,\cdots,X_n$ 下是 $(1-X_n,1]$ 上的均匀分布，$n=0,1,2,\cdots$，证明 $\{X_n,n=0,1,\cdots\}$ 的均值有遍历性。

17. 设 $\{\varepsilon_n,n=0,\pm 1,\cdots\}$ 为白噪声序列。令

$$X_n=\alpha X_{n-1}+\varepsilon_n, |\alpha|<1,\ n=\cdots,-1,0,1,\cdots.$$

则 $X_n=\sum\limits_{k=0}^{\infty}\alpha^k\varepsilon_{n-k}$，从而证明 $\{X_n,\ n=\cdots,-1,\ 0,\ 1,\cdots\}$ 为平稳序列。求出该序列的协方差函数。此序列是否具有遍历性?

18. 我们称一个随机过程 X 为平稳 Gauss-Markov 过程，如果 X 是平稳 Gauss 过程，并且具有 Markov 性，即对任意的 $s<t$，任意实数 x_t,x_s,x_u，有

$$P(X_t\leqslant x_t|X_s=x_s,X_u=x_u,u<s)=P(X_t\leqslant x_t|X_s=x_s).$$

试证明平稳 Gauss-Markov 过程的协方差函数 $R(\tau)$ 具有 $ce^{-a|\tau|}$ 这种形式，这儿 c 为常数。

19. 设 $\{X_n,n=0,1,\cdots\}$ 是平稳 Gauss-Markov 序列（即 18 题中的 t 取非负整数），均值为 0。证明其协方差函数 $R(h)$ 具有 $\sigma^2a^{|m|}$ 这种形式，其中 $|a|\leqslant 1$。

20. 设 $\{X(t)\}$ 为平稳过程。令 $Y(t)=X(t+a)-X(t-a)$。分别以 R_X；S_X 和 R_Y，S_Y 记随机过程 X 和 Y 的协方差函数和功率谱密度函数。证明

$$R_Y(\tau)=2R_X(\tau)-R_X(\tau+2a)-R_X(\tau-2a),$$

$$S_Y(\omega)=4S_X(\omega)\sin^2 a\omega.$$

21. 设平稳过程 X 的协方差函数 $R(\tau)=\sigma^2e^{-\tau^2}$。试研究其功率谱密度函数的性质。

22. 设平稳过程 $\{X(t)\}$ 的协方差函数 $R(\tau)=\dfrac{a^2}{2}\cos\omega\tau$

$+b^2e^{-a|\tau|}$，求功率谱密度函数 $S(\omega)$。

23．设 $\{X(t)\}$ 为 Gauss 平稳过程，均值为零，$R_X(\tau) = Ae^{-a|\tau|}\cos\beta\tau$。令 $Y(t)=X^2(t)$。验证 $R_Y(\tau)=A^2e^{-2a|\tau|}(1+\cos 2\beta\tau)$。并求出 $S_Y(\omega)$。

24．设 $\{X(t)\}$ 为 Gauss 平稳过程，均值为零，功率谱密度 $S(\omega)=\dfrac{1}{1+\omega^2}$。求 $X(t)$ 落在区间 $[0.5,1]$ 中的概率。

25．已知平稳过程 $\{X(t)\}$ 的功率谱密度为 $S(\omega)=\dfrac{\omega^2}{\omega^4+4\omega^2+3}$，求 $X(t)$ 的均方值。

26．设 $S(\omega)$ 是功率谱密度函数。证明 $\dfrac{d^2S(\omega)}{d\omega^2}$ 不可能是功率谱密度函数。

27．求下列协方差函数的功率谱密度函数

(1) $R(\tau)=\sigma^2e^{-a|\tau|}\cos b\tau$。

(2) $R(\tau)=\sigma^2e^{-a|\tau|}(\cos b\tau - ab^{-1}\sin b|\tau|)$，

(3) $R(\tau)=\sigma^2e^{-a|\tau|}(\cos b\tau + ab^{-1}\sin b|\tau|)$，

(4) $R(\tau)=\sigma^2e^{-a|\tau|}(1+a|\tau|-2a^2\tau^2+a^3|\tau|^3/3)$。

28．求下列功率谱密度对应的协方差函数

(1) $S(\omega)=\dfrac{\omega^2+64}{\omega^4+29\omega^2+100}$，

(2) $S(\omega)=\dfrac{1}{(1+\omega^2)^2}$，

(3) $S(\omega)=\sum\limits_{k=1}^{N}\dfrac{a_k}{\omega^2+b_k^2}$，$N$ 为固定的正整数。

(4) $S(\omega)=\begin{cases}a, & |\omega|\leqslant b,\\ 0, & |\omega|>b.\end{cases}$

(5) $S(\omega)=\begin{cases}0, & |\omega|<a \text{ 或 } |\omega|>2a,\\ b^2, & a\leqslant|\omega|\leqslant 2a.\end{cases}$

29．设 $\{\varepsilon_n, n=\cdots,-1,0,1\}$ 为白噪声序列。均值为 0，方差

为 σ^2。求下列序列的谱密度函数。

(1) $X_n=\varepsilon_n+a_1\varepsilon_{n-1}$, $n=\cdots,-1,0,1,\cdots$,

(2) $X_n=\sum\limits_{k=0}^{\infty}\beta_k\varepsilon_{n-k}$, 其中$\sum\limits_{k=0}^{\infty}\beta_k^2<\infty$。

30．由书中例 15 确定的平稳序列的功率谱密度是周期函数，试作出$(-\pi,\pi]$中谱密度函数的图形，并讨论当$|\rho|\to 1$时图形如何变化。

提示：分 $\rho>0$ 和 $\rho<0$ 讨论。

31．设$\{X_n, n=\cdots,-1,0,1,\cdots\}$为平稳序列，协方差函数为 $R(\tau)$。

(1) 求 X_{n+1}的形如 $\hat{X}_{n+1}^{(1)}=aX_n$ 的最小误差方差预报，这儿 a 是待定常数。

(2) 求 X_{n+1} 的形如 $\hat{X}_{n+1}^{(2)}=aX_n+bX_{n-1}$ 的最小均方误差预报，这儿 a 和 b 是待定常数。

(3) 上述两个预报 $\hat{X}_{n+1}^{(1)}$ 和 $\hat{X}_{n+1}^{(2)}$ 中，哪个预报的均方误差要小些？试用 $R(\tau)$表示它们的差。

(4) 求 X_{n+k} 的形如 $\hat{X}_{n+k}=aX_n+bX_{n+N}$, $1\leqslant k\leqslant N$ 的最小均方误差内插，这儿 a，b 为待定常数。

(5) 设 $Z_n=\sum\limits_{k=0}^{N}X_{n+k}$, 其中 N 为固定的正整数。求 Z_n 的形如 $\hat{Z}_n=aX_n+bX_{n+N}$ 的最小均方误差预报，其中 a，b 为待定常数。

32．设平稳序列$\{X_n,n=\cdots,-1,0,1,\cdots\}$的均值为零，协方差函数为$R(h)=\rho^{|h|}$, $|\rho|<1$, $h=0,1,\cdots$求X_{n+1}根据$\{X_k,k\leqslant n\}$的线性最佳预报 $\hat{X}_{n+1}$。

以下设 $\{\varepsilon_n,n=\cdots,-1, 0, 1,\cdots\}$是均值为0，方差为1的白噪声序列。

33．证明没有一个平稳序列$\{X_n,n=0,\pm 1,\cdots\}$能满足$X_n=X_{n-1}+\varepsilon_n$, $n=0$, $\pm 1,\cdots$。

34．证明如下两个滑动平均序列

$$X_n = \varepsilon_n + \alpha\varepsilon_{n-1},$$
$$Y_n = \varepsilon_n + \frac{1}{\alpha}\varepsilon_{n-1}.$$
有相同的相关函数。

35．设 $\{X_n, n=0, \pm 1, \cdots\}$ 为 $AR(p)$ 模型：
$$X_n = \alpha_1 X_{n-1} + \cdots + \alpha_p X_{n-p} + \varepsilon_n, n = \cdots, -1, 0, 1, \cdots.$$
试导出（Yule-Walker）方程：
$$R(h) = \alpha_1 R(h-1) + \cdots + \alpha_p R(h-p), \quad h > 0.$$
提示：$AR(p)$ 模型两边同乘以 X_{n-h}，然后取期望。

36．考虑 $AR(p)$ 模型：
$$X_n = \alpha_1 X_{n-1} + \cdots + \alpha_p X_{n-p} + \varepsilon_n, n = \cdots, -1, 0, 1, \cdots.$$
假定 $1 - \alpha_1 z - \cdots - \alpha_p z^p$ 的根都在单位圆外，求功率谱密度函数。

37．考虑如下 $AR(2)$ 模型：

(1) $X_n = 0.5X_{n-1} + 0.3X_{n-2} + \varepsilon_n$，

(2) $X_n = 0.5X_{n-1} - 0.3X_{n-2} + \varepsilon_n$.

试用 Yule-Walker 方程导出协方差函数，证明它们的谱密度函数 $S(\omega)$ 是周期函数，并作出 $S(\omega)$ 在 $(-\pi, \pi)$ 上的图形。

38．求下列自回归模型的协方差函数和相关函数：

(1) $X_n = 0.8X_{n-1} + \varepsilon_n$，

(2) $X_n = 0.4X_{n-1} + \varepsilon_n$，

(3) $X_n = -0.5X_{n-1} + \varepsilon_n$.

39．求下列滑动平均模型的协方差函数和相关函数：

(1) $X_n = \varepsilon_n - 0.5\varepsilon_{n-1} - 0.5\varepsilon_{n-2}$，

(2) $X_n = \varepsilon_n + 0.6\varepsilon_{n-1} - 0.2\varepsilon_{n-2} - 0.1\varepsilon_{n-3}$.

40．设
$$X_n = \varepsilon_n + \beta[\varepsilon_{n-1} + \gamma\varepsilon_{n-2} + \gamma^2\varepsilon_{n-3} + \cdots],$$
这儿 β 和 γ 为常数，$|\gamma| < 1$，记 $\alpha = \gamma - \beta$，$|\alpha| < 1$。求 X_{n+1} 根据 $\{X_k, k \leq n\}$ 的线性最佳预报。

41．考虑 $AR(2)$ 模型：

$$X_n = 1.8X_{n-1} + 0.8X_{n-2} + \varepsilon_n,$$

求一步预报及 $l>1$ 步预报 $\hat{X}_{n+1|n}$，$\hat{X}_{n+l|n}$。

42．考虑 $AR(2)$模型：

$$X_n = X_{n-1} - 0.25X_{n-2} + \varepsilon_n,$$

求 $\hat{X}_{n+l|n}$ 及 $E(X_{n+l} - \hat{X}_{n+l|n})^2$。

附　　录　　A

定理 4.5 的证明。

充分性。设 $\hat{X}^*$ 是 $\mathscr{H}$ 中一个预报，对每个 $\mathscr{H}$ 中的预报 U 有 $E[(X-\hat{X}^*)U]=0$。我们将证明 $\hat{X}^*$ 有最小的均方误差。设 $\hat{X}$ 为任一允许的预报(即 $\hat{X}\in\mathscr{H}$)，我们有

$$E(X-\hat{X})^2 = E(X-\hat{X}^*)^2 + 2E(X-\hat{X}^*)(\hat{X}^*-\hat{X}) + E(\hat{X}^*-\hat{X})^2.$$

令 $U=\hat{X}^*-\hat{X}$，它是 $\hat{X}^*$ 和 $\hat{X}$ 的线性组合，由定理条件知 $U\in\mathscr{H}$。因此由(4.42)知

$$E(X-\hat{X}^*)(\hat{X}^*-\hat{X}) = E[(X-\hat{X}^*)U]=0,$$

从而

$$E(X-\hat{X})^2 = E(X-\hat{X}^*)^2 + E(\hat{X}^*-\hat{X})^2 \geqslant E(X-\hat{X}^*)^2.$$

即 $\hat{X}^*$ 有最小的均方误差。

必要性。设$\hat{X}^*\in\mathscr{H}$为最佳预报。要证对每个$U\in\mathscr{H}$,(4.42)式成立。用反证法，若否，设对某个$U\in\mathscr{H}$有 $E[(X-\hat{X}^*)U]=\alpha\neq 0$。令

$$\hat{X} = \hat{X}^* + (\alpha/EU^2)U,$$

即 $\hat{X}-\hat{X}^* = (\alpha/EU^2)U$。则

$$\begin{aligned}E(X-\hat{X})^2 &= E[\{(X-\hat{X}^*)-(\hat{X}-\hat{X}^*)\}^2]\\ &= E(X-\hat{X}^*)^2 - 2E(X-\hat{X}^*)(\hat{X}-\hat{X}^*) + E(\hat{X}-\hat{X}^*)^2\\ &= E(X-\hat{X}^*)^2 - (2\alpha/EU^2)E[(X-\hat{X}^*)U] + (\alpha^2/(EU^2)^2)\\ &\quad\cdot EU^2 = E(X-\hat{X}^*)^2 - (\alpha^2/EU^2) < E(X-\hat{X}^*)^2.\end{aligned}$$

即 $\hat{X}$ 的均方误差小于 $\hat{X}^*$ 的均方误差，因而$\hat{X}^*$ 不能是最佳预报，与条件矛盾。必要性证毕。

附 录 B

设 $\mathscr{H}$ 为由 $1, X_1, \cdots, X_n$ 生成的线性空间，即 $\mathscr{H}$ 中的元素是 $1, X_1, \cdots, X_n$的线性组合。X为任一二阶矩存在的随机变量。若 X 的预报类仅限于 $\mathscr{H}$中，即只能是 $1, X_1, \cdots, X_n$ 的线性组合，记$Y=(X_1, \cdots, X_n)$，$\hat{E}(X|Y)$为X在 $\mathscr{H}$ 上的投影，由定理 4.5 即知 $\hat{E}(X|Y)$就是X的基于 1，$X_1, \cdots, X_n$的线性最佳预报。且有如下的显式表达式：

$$\hat{E}(X|Y)=EX+\mathrm{Cov}(X,Y)(\mathrm{Var}Y)^{-1}(Y'-EY'), \quad (B.1)$$

其中 Y' 表示向量 Y 的转置，且

$$\mathrm{Cov}(X,Y)=(\mathrm{Cov}(X,X_1),\mathrm{Cov}(X_1,X_2), \cdots, \mathrm{Cov}(X,X_n)),$$

$$\mathrm{Var}(Y)=\begin{pmatrix} \mathrm{Var}(X_1)\mathrm{Cov}(X_1,X_2)\cdots\mathrm{Cov}(X_1,X_n) \\ \mathrm{Cov}(X_1,X_2) \ \ \mathrm{Var}(X_2)\cdots\mathrm{Cov}(X_2,X_n) \\ \vdots \\ \mathrm{Cov}(X_1,X_n) \ \ \mathrm{Cov}(X_2,X_n)\cdots\mathrm{Var}(X_n) \end{pmatrix}.$$

证 由最小误差方差预报的定义，我们就是求 $\mathscr{H}$ 中的元素 $\hat{X}$，使

$$E(X-\hat{X})^2 \quad (B.2)$$

达到极小。令 $\hat{X}=a_0+AY'$，$A=(a_1, \cdots, a_n)$，则我们就是要选取 a_0 和 A 使$E(X-a_0-AY')^2$达到极小，令

$$b=a_0-EX+AEY', \quad (B.3)$$

其中 $EY'=(EX_1, \cdots, EX_n)'$。则$(B.2)$可改写为

$$\begin{aligned} E(X-\hat{X})^2 &= E[X-EX-b-A(Y'-EY')]^2 \\ &= \mathrm{Var}(X)+b^2+A[\mathrm{Var}(Y))]A' \\ &\quad -\mathrm{Cov}(X,Y')A'-A\mathrm{Cov}(Y'X) \\ &= b^2+[A-\mathrm{Cov}(X,Y)(\mathrm{Var}(Y))^{-1}](\mathrm{Var}(Y)) \end{aligned}$$

$$\cdot[A-\mathrm{Cov}(X,Y)(\mathrm{Var}(Y))^{-1}]'+[\mathrm{Var}(X)-\mathrm{Cov}(X,Y)(\mathrm{Var}(Y))^{-1}\mathrm{Cov}(Y',X)]. \tag{B.4}$$

注意(B.4)中右边前二项都是非负的，而第三项与 b，A 无关。因此为使(B.4)达到极小，唯一的解就是选取 b 和 A，使得前二项都为零。即

$$b=0, \tag{B.5}$$

$$A=\mathrm{Cov}(X,Y)(\mathrm{Var}(Y))^{-1}. \tag{B.6}$$

把(B.5)和(B.6)两式代入(B.3)，得

$$a_0=EX-\mathrm{Cov}(X,Y)(\mathrm{Var}(Y))^{-1}EY'. \tag{B.7}$$

由(B.6)和(B.7)即得线性最佳预报为

$$\begin{aligned}\hat{E}(X|Y)&=EX-\mathrm{Cov}(X,Y)(\mathrm{Var}(Y))^{-1}EY'\\&\quad+\mathrm{Cov}(X,Y')(\mathrm{Var}Y)^{-1}Y'\\&=EX+\mathrm{Cov}(X,Y)(\mathrm{Var}(Y))^{-1}(Y'-EY').\end{aligned}$$

第五章 Brown运动

英国植物学家布朗(Brown)在1827年观察到悬浮在液体中的小颗粒粒子的运动是非常不规则的，这种运动称为Brown运动。自Brown运动发现以来，这个过程已大量地在数理统计、分析贮存商品的价格水平、量子力学、通信理论、生物学、管理科学等领域得到广泛应用。下面我们对这一过程作一简单介绍以作为进一步学习的基础。

5.1 定　　义

作为一种物理现象，Brown运动是英国植物学家Robert Brown在1827年发现的，人们最早的想法是如何阐述这种现象。关于这个现象的数学解释是爱因斯坦(Einstein)在1905年首先从物理定律中导出的。物理理论由Ornstein, Uhlenbeck等进一步发展完善。但在数学理论方面的发展要慢一点，这是因为关于这个模型如何用数学来精确描述存在困难。直到1918年维纳(Wiener)在他的博士论文及以后的一系列论文中才得到了Brown运动精确的数学公式。所以Brown运动又称为Wiener过程。下面我们从几个不同的角度来导出布朗运动的性质，从而给出Brown运动数学上的定义。

设$X(t)$表示一个粒子作Brown运动中的x方向分量，x_0为粒子在时刻t_0的位置，即$X(t_0)=x_0$。设$p(x,t|x_0)$表示在给定$X(t_0)=x_0$的条件下$X(t+t_0)$的条件概率密度。我们假设所给的转移概率是平稳的，从而$p(x,t|x_0)$不依赖于起始时刻t_0。因为$p(x,t|x_0)$是x的密度函数，故

$$p(x,t|x_0)\geqslant 0, \quad \int_{-\infty}^{\infty} p(x,t|x_0)\,dx = 1. \tag{5.1}$$

进一步，我们要求对充分小的 t，$X(t+t_0)$ 与 $X(t_0)=x_0$ 非常接近，即

$$\lim_{t\to 0} p(x,t|x_0) = 0, \quad x\neq x_0. \tag{5.2}$$

由物理原理，爱因斯坦证明了 $p(x,t|x_0)$ 必然满足偏微分方程：

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2}. \tag{5.3}$$

上述方程称为扩散方程，D 称为扩散系数。小颗粒子进行Brown运动是由于受到大量分子的碰撞。每次碰撞的影响是微不足道的，但大量分子碰撞的迭加则产生了可观测的运动。D 的估计根据公式 $D=2RT/Nf$ 来确定，其中 R 为气体（液体）常数，T 为温度，N 为 Avogadro 数，f 为摩擦系数。选择适当的单位，我们可以取$D=\frac{1}{2}$。于是我们可以直接验证

$$p(x,\ t|x_0) = \frac{1}{\sqrt{2\pi t}}\exp\left\{-\frac{1}{2t}(x-x_0)^2\right\}. \tag{5.4}$$

是(5.3)的解。事实上，这是在边界条件(5.1)和(5.2)之下的唯一解（详细证明已超出本课程范围）。

爱因斯坦导出的(5.3)式也可以根据简单随机游动逼近的方式导出。考虑对称随机游动$\left(\text{即 } P(X_n=-1)=P(X_n=1)=\frac{1}{2}\right)$，其中 X_n 表示粒子在第 n 时刻移动的方向，$S_n=X_1+\cdots+X_n$ 表示在时刻 n 粒子的位置。以 $p_k(n)$ 表示时刻 n 时粒子处于位置 k 的概率，则由 Chapman-Kolmogorov 方程有

$$p_k(n+1) = \frac{1}{2}p_{k+1}(n) + \frac{1}{2}p_{k-1}(n),$$

这方程也可以改写为

$$p_k(n+1)-p_k(n)=\frac{1}{2}[p_{k+1}(n)-2p_k(n)+p_{k-1}(n)]. \quad (5.5)$$

该式左边是时间的一阶差分，而右边是位置变量的二阶差分。通过适当的极限过程让转移时间趋于 0，同时让步长适当收缩到 0，我们可以由(5.5)得(5.3)式。特别，设转移时间间隔为 Δt，步长为 Δx，则(5.5)式可写为

$$\frac{p_{k\Delta x}((n+1)\Delta t)-p_{k\Delta x}(n\Delta t)}{\Delta t}$$

$$=\frac{1}{2}\frac{p_{(k+1)\Delta x}(n\Delta t)-2p_{k\Delta x}(n\Delta t)+p_{(k-1)\Delta x}(n\Delta t)}{(\Delta x)^2}\cdot\frac{(\Delta x)^2}{\Delta t}. \quad (5.6)$$

然后令 $\Delta t\to 0$，$\Delta x\to 0$ 而保持 $\Delta t=(\Delta x)^2$。再令 $n\to\infty$，$k\to\infty$ 使 $k\Delta x\to x$，$n\Delta t\to t$，则 $p_{k\Delta t}(n\Delta x)\to p(x,t|x_0)$。由(5.6)即得(5.3)式。

当然上面这个极限过程是直观的，真正要把它精确化需要相当细致的分析，我们不能在这儿叙述了。

关于用简单随机游动逼近 Brown 运动的方法除了上述 Chapman-Kolmogorov 方程外，还可以用中心极限定理的方法。设 $X(t)$ 表示粒子在时刻 t 时的位置。粒子转移时间间隔和步长分别为 Δt 和 Δx，X_k 表示在时刻 $k\Delta t$ 时转移方向，则

$$X(t)=\Delta x(X_1+X_2+\cdots+X_{[t/\Delta t]}). \quad (5.7)$$

其中 $[x]$ 表示 x 的整数部分。由于 $EX_i=0$，$\mathrm{Var}(X_i)=1$，故在(5.7)中，$EX(t)=0$，$\mathrm{Var}(X(t))=(\Delta x)^2\left[\frac{t}{\Delta t}\right]$，现令 Δx，$\Delta t\to 0$，但趋于 0 的方式应保持方差不能趋于 0 或无穷，即要求 $(\Delta x)^2=c^2\Delta t$（因为若 $\Delta x=\Delta t$，则 $\mathrm{Var}(X(t))\to 0$，从而由概率论知识得 $X(t)$ 几乎等于 0。如果 $\Delta t=(\Delta x)^3$，则 $\mathrm{Var}(X(t))\to\infty$，这也是不合实际情况的。因为由物理上粒子运动的连续性，不可能在很短时间内远离出发点。因此合理的假设只能是 $(\Delta x)^2=c^2\Delta t$，其中 $c>0$）。再令 $\Delta t\to 0$，则

$$\mathrm{Var}(X(t)) \to c^2 t.$$

由中心极限定理，可以得到

$$X(t) \sim N(0, c^2 t).$$

此外，由于在不相交时间间隔内的随机游动是独立的。因此我们可以知道 Brown 运动有独立增量。最后，由于在任何一段时间内随机游动位置变化的分布只依赖于区间的长度，故 Brown 运动应有平稳增量。综上所述，我们给出 Brown 运动的定义。

定义 5.1 随机过程 $\{X(t), t \geqslant 0\}$ 称为 Brown 运动，如果它满足如下三个条件

(i) $X(0) = 0$,

(ii) 随机过程 X 有平稳独立增量，

(iii) 对每个 $t > 0$, $X(t) \sim N(0, c^2 t)$.

如果 $c = 1$，我们称为标准Brown运动，如果 $c \neq 1$，则可考虑 $\{X(t)/c, t \geqslant 0\}$ 则 $\{X(t)/c, t \geqslant 0\}$ 就是标准Brown运动，故不失一般性，可以只考虑标准Brown运动的性质。此外，我们也经常考虑 $(-\infty, \infty)$ 上的Brown运动，此时只要把上面条件 (iii) 改为

$$X(t) \sim N(0, c^2|t|), \quad -\infty < t < \infty.$$

5.2 Brown 运动的性质

下面我们研究 Brown 运动的轨道性质和联合分布等概率性质。

首先Brown运动的几乎每条样本轨道是连续的（见图5.1）。这可以从物理上直观得到，也可以从Brown运动是随机游动极限的解释上得到启发。这一点在数学上是可以严格证明的。虽然如此，但 $X(t)$ 的样本轨道不是通常我们所见到的函数，而是一个几乎点点都没有导数的函数。其物理意义是粒子在每一瞬间，都可能受到介质中分子的碰撞，碰撞后的方向是任意的，而速度为无穷。

从数学上看，可以粗略地由下面分析看出。若 $X(t)$ 在 t_0 有导数，则 $\frac{X(t_0+\Delta t)-X(t_0)}{\Delta t}\to X'(t_0)\ (\Delta t\to 0)$，但由 Brown 运动的平稳性和正态性知

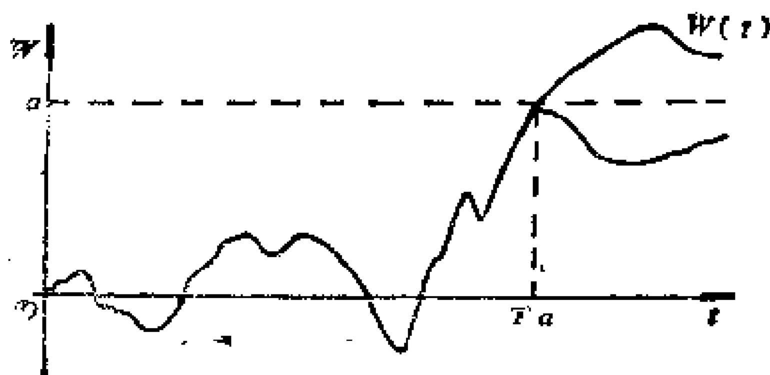


图 5.1

$$X(t_0+\Delta t)-X(t_0)\sim N(0,\Delta t).$$

因此对任给的正数 M，有

$$\begin{aligned}&P\left(\left|\frac{X(t_0+\Delta t)-X(t_0)}{\Delta t}\right|\leqslant M\right)\\&=P\left(\left|\frac{X(\Delta t)}{\sqrt{\Delta t}}\right|\leqslant M\sqrt{\Delta t}\right)\\&=\Phi(M\sqrt{\Delta t})-\Phi(-M\sqrt{\Delta t})\to 0\quad(\Delta t\to 0).\end{aligned}$$

此式表示，Brown 运动在任一点 t_0 的导数有限的概率为 0，换句话说，对几乎每条样本轨道上的任意一点 t_0，其导数都不存在。

下面我们进一步研究 Brown 运动的概率特性。由定义，如果 $0\leqslant s$，$0\leqslant t$，则 $X(t+s)-X(s)$ 与时刻 v，$0\leqslant v<s$ 的位置 $X(v)$ 相互独立，因此

$$\begin{aligned}&P(X(t+s)\leqslant a\mid X(s)=x,X(v)=x_v,0\leqslant v<s)\\&=P(X(t+s)-X(s)\leqslant a-x\mid X(s)=x,X(v)=x_v,0\leqslant v<s)\\&=P(X(t+s)-X(s)\leqslant a-x)\\&=P(X(t+s)\leqslant a\mid X(s)=x).\end{aligned}\tag{5.8}$$

上面第二个等式由 Brown 运动的独立增量得到，最后一个等式

由条件概率等式得到。由此可知，在给定现在状态 $X(s)$ 的条件下，过去 $X(v)$，$0\leqslant v<s$，与将来 $X(t+s)$，$t>0$，独立，即由定义 3.7 知它是一个齐次 Markov 过程。

其次，因为 $X(t)$ 是均值为零，方差为 t 的正态分布，故它的密度函数为

$$f_t(x)=\frac{1}{\sqrt{2\pi t}}e^{-x^2/2t},$$

任给 n 个时刻 $0<t_1<t_2<\cdots<t_n$，以 $f_{t_1,\cdots,t_n}(x_1,\cdots,x_n)$ 记这 n 个时刻的位置 $(X(t_1),\cdots,X(t_n))$ 的联合分布密度函数，则我们有

定理 5.1

$$f_{t_1,\cdots,t_n}(x_1,\cdots,x_n)=f_{t_1}(x_1)f_{t_2-t_1}(x_2-x_1)\cdots f_{t_n-t_{n-1}}(x_n-x_{n-1}). \tag{5.9}$$

证 对 n 用归纳法。$n=2$ 时由 (5.8) 及 Brown 运动的平稳性有

$$P(X_{t_2}\leqslant x_2|X_{t_1}=x_1)=P(X_{t_2-t_1}\leqslant x_2-x_1),$$

故给定 $X_{t_1}=x_1$ 时 X_{t_2} 的条件密度函数为

$$f_{t_2}(x_2|X_{t_1}=x_1)=f_{t_2-t_1}(x_2-x_1),$$

所以由条件密度性质知

$$\begin{aligned}f_{t_1,t_2}(x_1,x_2)&=f_{t_2}(x_2|X_{t_1}=x_1)f_{t_1}(x_1)\\&=f_{t_1}(x_1)f_{t_2-t_1}(x_2-x_1).\end{aligned}$$

设 $n=k$ 时(5.9)式成立，当 $n=k+1$ 时，由(5.8)式及平稳性有

$$\begin{aligned}&P(X_{t_{k+1}}\leqslant x_{k+1}|X_{t_k}=x_k,\quad X_{t_j}=x_j,\quad 1\leqslant j\leqslant k-1)\\&=P(X_{t_{k+1}}-X_{t_k}\leqslant x_{k+1}-x_k|X_{t_k}=x_k)\\&=P(X_{t_{k+1}}-X_{t_k}\leqslant x_{k+1}-x_k).\end{aligned}$$

因此给定 $X_{t_k}=x_k$，$X_{t_j}=x_j$，$1\leqslant j\leqslant k-1$ 时 $X_{t_{k+1}}$ 的条件密度为

$$f_{t_{k+1}}(x_{t_{k+1}}|X_{t_j}=x_j,\ 1\leqslant j\leqslant k)=f_{t_{k+1}-t_k}(x_{k+1}-x_k).$$

由归纳假设及条件密度性质得

$$f_{t_1,\cdots,t_{k+1}}(x_1,\cdots,x_{k+1}) = f_{t_1,\cdots,t_k}(x_1,\cdots,x_k) f_{t_{k+1}}(x_{t_{k+1}} |$$

$$X_{t_j} = x_j, 1 \leqslant j \leqslant k) = f_{t_1,\cdots,t_k}(x_1,\cdots,x_k) f_{t_{k+1}-t_k}(x_{k+1} - x_k)$$

$$= f_{t_1}(x_1) f_{t_2-t_1}(x_2 - x_1) \cdots f_{t_{k+1}-t_k}(x_{k+1} - x_k).$$

由归纳法知(5.9)式对一切正整数 n 成立。

由定理 5.1，原则上我们可以计算所需求的任一概率。例如，我们可以求出在给定 $X(t)=B$ 时，$X(s)$，$s<t$，的条件密度：

$$f_{s|t}(x|B) = \frac{f_{s,t}(x,B)}{f_t(B)} = \frac{f_s(x) f_{t-s}(B-x)}{f_t(B)}$$

$$= \frac{\sqrt{2\pi t}}{\sqrt{2\pi s} \cdot \sqrt{2\pi(t-s)}} \exp\left\{-\frac{x^2}{2s} - \frac{(B-x)^2}{2(t-s)} + \frac{B^2}{2t}\right\}$$

$$= \sqrt{\frac{t}{2\pi s(t-s)}} \exp\left\{-\frac{t(x-Bs/t)^2}{2s(t-s)}\right\}.$$

即服从均值为 Bs/t，方差为 $s(t-s)/t$ 的正态分布。

有意思的是，在给定 $X(t)=B$ 时 $X(s)$，$s<t$，的条件方差不依赖于 B。如果令 $s/t=\alpha$，$0<\alpha<1$，则给定 $X(t)$ 之下 $X(s)$ 的条件分布是均值为 $\alpha X(t)$ 和方差为 $\alpha(1-\alpha)t$ 的正态分布。

由定理 5.2 知对任意 n 个时刻的位置 $(X(t_1),\cdots,X(t_n))$ 服从 n 维正态分布，从而 Brown 运动是一个 Gauss 过程。我们知道，Gauss 过程的概率性质由均值函数和协方差函数完全确定。对 Brown 运动来说，其均值和协方差函数为(设 $s\leqslant t$)

$$EX(t) = 0,$$

$$\mathrm{Cov}(X(s),X(t)) = \mathrm{Cov}(X(s), X(s)+X(t)-X(s))$$

$$= \mathrm{Cov}(X(s),X(s)) + \mathrm{Cov}(X(s),X(t)-X(s))$$

$$= s.$$

当 $t>s$ 时，$X(s)$ 和 $X(t)$ 的协方差为 t。故

$$\mathrm{Cov}(X(s),X(t)) = \min(s,t).$$

由 Brown 运动，我们可以定义另一个非常重要的过程——Brown 桥(Brown Bridge)过程。

定义 5.2 设 $\{W(t):t\geqslant 0\}$ 为 Brown 运动。令

$$B(t)=W(t)-tW(1),\ 0\leqslant t\leqslant 1,$$

则随机过程 $B=\{B(t):0\leqslant t\leqslant 1\}$ 称为Brown桥过程。

由于 Brown 运动是 Gauss 过程，故 Brown 桥过程也是 Gauss 过程，其 n 维分布由均值函数和协方差函数完全确定。容易算得，对 $\forall 0\leqslant s\leqslant t\leqslant 1$，

$$EB(t)=0,$$

$$\begin{aligned} EB(s)B(t) &= E[(W(s)-sW(1))(W(t)-tW(1))] \\ &= E(W(s)W(t)-tW(s)W(1)-sW(t)W(1)+tsW^2(1)) \\ &= s-ts-ts+ts=s(1-t). \end{aligned} \tag{5.10}$$

此外，由 Brown 桥定义知 $B(0)=0$和 $B(1)=0$，即该运动的两头都是固定的，就象桥一样，这就是 Brown 桥名字的由来，由前面讨论知，在给定 $W(1)=0$ 的条件下，$W(s)$，$0\leqslant s\leqslant 1$ 的条件分布就是 Brown 桥过程.

下面将看到，Brown 桥在研究经验分布函数中起着十分有力的作用.

除了上述性质外，Brown 运动还具有平移不变性和刻度不变性。即如果 $\{W(t):t\geqslant 0\}$ 为 Brown 运动，a 和 c 为两个正常数，则 $\{W(t)-W(a):t\geqslant a\}$ 和 $\{W(ct)/\sqrt{c}:t\geqslant 0\}$也是 Brown 运动(见习题)。

还有两个与 Brown 运动相关的随机变量在应用中经常遇到。一个是首达时 T_a (见图 5.1)，它是布朗运动首次到达 a 的时刻。现在我们来求它的分布. 先设 $a>0$. 可以分两种情况来考虑。一是 $T_a\leqslant t$，即在时刻 t 以前 Brown 运动已到达 a。另一种是 $T_a>t$，运动没有到达或经过 a 点. 由全概率公式得

$$\begin{aligned} P(W(t)\geqslant a) &= P(W(t)\geqslant a\,|\,T_a\leqslant t)P(T_a\leqslant t) \\ &\quad + P(W(t)\geqslant a\,|\,T_a>t)P(T_a>t). \end{aligned} \tag{5.11}$$

上式第二个概率由 T_a 的定义知必为零。故只要计算第一个概率。设 $T_a\leqslant t$，即在 $[0, t]$ 中的某一时刻，Brown运动首次过 a。

(见图 5.2) 构造新的 Brown 运动 $\widetilde{W}=\{\widetilde{W}(t);\ t\geqslant T_a\}$，其中 $\widetilde{W}(t)=a-(W(t)-a)$，$t\geqslant T_a$。则由对称性，当 $t\geqslant T_a$ 时

$$P(\widetilde{W}(t)\geqslant a)=P(\widetilde{W}(t)\leqslant a)=\frac{1}{2},$$

即当 $t\geqslant T_a$ 时

$$P(W(t)\geqslant a)=P(W(t)\leqslant a)=\frac{1}{2}.$$

写成条件概率的形式即为

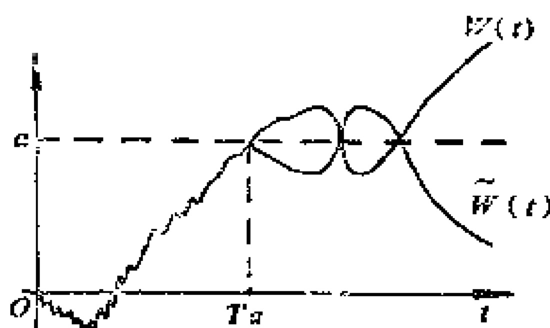


图 5.2

$$P(W(t)\geqslant a\mid T_a\leqslant t)=P(W(t)\leqslant a\mid T_a\leqslant t)=\frac{1}{2},\qquad(5.12)$$

代入(5.11)式，即得

$$P(W(t)\geqslant a)=\frac{1}{2}P(T_a\leqslant t),$$

即当 $a>0$ 时，

$$P(T_a\leqslant t)=2P(W(t)\geqslant a)=\frac{2}{\sqrt{2\pi t}}\int_a^{\infty}e^{-x^2/2t}dx$$

$$=\sqrt{\frac{2}{\pi}}\int_{a/\sqrt{t}}^{\infty}e^{-y^2/2}dy.\qquad(5.13)$$

由上式可推知

$$P(T_a<\infty)=\lim_{t\to\infty}P(T_a\leqslant t)=\sqrt{\frac{2}{\pi}}\int_0^{\infty}e^{-y^2/2}dy=1.$$

即对几乎每条样本轨道，不管 a 多大，总在有限时刻首次过 a。

由 (5.13) 式我们可以计算初次过 a 所需的平均时间 ET_a。由 (5.13)，T_a 的密度函数为

$$f_{T_a}(t)=\frac{a}{\sqrt{2\pi t^3}}e^{-a^2/2t},\quad t>0。\tag{5.14}$$

因此

$$ET_a=\int_0^\infty t\cdot f_{T_a}(t)\,dt=\int_0^\infty\frac{ta}{\sqrt{2\pi t^3}}e^{-a^2/2t}\,dt$$
$$=\int_0^\infty\frac{a}{\sqrt{2\pi t}}e^{-a^2/2t}\,dt。$$

因为当 $t\to\infty$ 时 $\frac{a}{\sqrt{2\pi t}}e^{-a^2/2t}\sim\frac{a}{\sqrt{2\pi t}}$，而

$$\int_1^\infty\frac{dt}{\sqrt{t}}=\infty。$$

因此

$$ET_a=\infty。$$

上式说明虽然 Brown 运动的几乎所有样本轨道初次过 a 的时刻都有限，即几乎每条轨道都能在有限时间过 a 点，但是过 a 点的平均时间却是无穷。

如果 $a<0$，由对称性，其分布与 T_{-a} 相同，因此我们有

$$P(T_a\leqslant t)=\sqrt{\frac{2}{\pi}}\int_{|a|/\sqrt{t}}^\infty e^{-y^2/2}\,dy。\tag{5.15}$$

另一个重要的随机变量是 Brown 运动在 $[0,t]$ 上的最大值 $\max\limits_{0\leqslant s\leqslant t}W(s)$。由于 Brown 运动的几乎所有样本是连续的，所以集合 $(\max\limits_{0\leqslant s\leqslant t}W(s)\geqslant a)=(T_a\leqslant t)$。注意到 $W(0)=0$，故 $\max\limits_{0\leqslant s\leqslant t}W(s)\geqslant 0$，故当 $a\geqslant 0$ 时

$$P(\max_{0\leqslant s\leqslant t}W(s)\geqslant a)=P(T_a\leqslant t)=\sqrt{\frac{2}{\pi}}\int_{a/\sqrt{t}}^\infty e^{-y^2/2}\,dy。\tag{5.16}$$

故随机变量 $M(t)\triangleq\max\limits_{0\leqslant s\leqslant t}W(s)$ 的密度函数 $f_{M(t)}(a)$ 为

$$f_{M(t)}(a)=\begin{cases}\dfrac{2}{\sqrt{2\pi t}}\exp\{-a^2/2t\}, & a\geqslant 0,\\ 0, & a<0.\end{cases}$$

5.3 随机积分和随机微分方程

Brown运动是随机过程中最重要的过程之一，随着现代概率论，随机微分方程的迅速发展，越来越显示出 Brown 运动的重要性，它的重要性大体表现在两个方面，一是通过对Brown运动及其函数的研究，可以得到一系列在经典场合下难以得到的结果，同时也增加了新的内容。二是关于对 Brown 运动的随机积分，例如 *Ito* 微分公式，开创并发展了关于随机微分方程的研究。这方面的研究给物理学等提供了更有力的工具，所以在统计物理、通讯、经济等各个领域都得到了广泛的应用。可以夸张一点地说，不了解 Brown 运动就不可能真正了解现代概率论。这一节我们将对随机积分及随机微分方程作一简单介绍。

5.3.1 积分

根据实际需要，我们经常要计算关于随机过程的微分和积分。例如在第四章，我们已经遇到宽平稳过程的积分问题：

$$\overline{X}^T=\frac{1}{2T}\int_{-T}^{T}X(t)\,\mathrm{d}t$$

这儿 $\overline{X}^T$ 是随机变量（因为 $X(t)=X(t,\omega)$ 是随机变量）。这与统计学中用样本均值 $\overline{X}$ 来估计总体均值 μ 相似。如同通常积分一样，对被积过程也要有一定的限制。为此我们需要定义如下随机过程。

定义 5.3 设 $X=\{X(t),t\in T\}$ 是一随机过程，若对每个 $t\in T$，都有

$$EX^2(t)<\infty,$$

则称 $\{X(t), t\in T\}$ 为二阶矩过程。

由定义及 Schwarz 不等式可知，二阶矩过程 X 的均值函数 $m_X(t)$ 和协方差函数 $r_X(s,t)$ 都存在。这是一个很大的过程类，许多我们经常遇到的过程，例如宽平稳过程，Gauss 过程，Brown 运动和 Brown 桥过程都是二阶矩过程。下面我们还设二阶矩过程满足如下要求：

(i) T 是一个区间或一条直线，

(ii) 对每个固定的 $\omega\in\Omega$，X 的轨道 $X(t)$ 是分段连续的，

(iii) $m_X(t)$ 和 $r_X(s,t)$ 是区间 T 上的连续函数。

设 $f(t)$ 是 T 上的连续函数，对每个 $\omega\in\Omega$，由上述条件 (i) 和(ii)，我们可以定义二阶矩过程 X 的如下积分：

$$\int_a^b f(t)X(t)\mathrm{d}t$$

其中 $[a,b]\subseteq T$。由于 $X(t)$ 与 ω 有关，所以上述积分表示一个随机变量。我们可以计算它的期望和方差，由于期望和积分号可以交换，故

$$E\int_a^b f(t)X(t)\mathrm{d}t=\int_a^b f(t)EX(t)\mathrm{d}t=\int_a^b f(t)m_X(t)\mathrm{d}t$$

$$\begin{aligned}E\left(\int_a^b f(t)X(t)\mathrm{d}t\right)^2&=E\left[\int_a^b f(t)X(t)\mathrm{d}t\int_a^b f(s)X(s)\mathrm{d}s\right]\\&=\int_a^b f(t)\left[\int_a^b f(s)E(X(t)X(s))\mathrm{d}s\right]\mathrm{d}t.\end{aligned}$$

因而

$$\mathrm{Var}\left(\int_a^b f(t)X(t)\mathrm{d}t\right)=\int_a^b f(t)\left(\int_a^b f(s)r_X(s,t)\mathrm{d}s\right)\mathrm{d}t.$$

如果 $g(t)$ 是 $[c,d]$ 上的连续函数，我们进而能得到

$$\begin{aligned}&E\left(\int_a^b f(t)X(t)\mathrm{d}t\int_c^d g(t)X(t)\mathrm{d}t\right)\\&=\int_a^b f(t)\left(\int_c^d g(s)EX(t)X(s)\mathrm{d}s\right)\mathrm{d}t,\end{aligned}\tag{5.17}$$

$$\mathrm{Cov}\left(\int_a^b f(t)X(t)\mathrm{d}t,\int_c^d g(t)X(t)\mathrm{d}t\right)$$

$$=\int_a^b f(t)\left(\int_c^d g(s)r_X(s,t)\,\mathrm{d}s\right)\mathrm{d}t. \tag{5.18}$$

如果X是Gauss过程，则$\int_a^b f(t)X(t)\,\mathrm{d}t$是正态随机变量。为看出这一点，注意到积分可以用如下的有限和来逼近：

$$\frac{b-a}{n}\sum_{k=1}^{n} f\left(a+\frac{b-a}{n}k\right)X\left(a+\frac{b-a}{n}k\right).$$

由正态随机变量线性组合仍是正态随机变量这一事实知，上式是一个正态随机变量，而且当 $n\to\infty$ 时收敛。可以证明，其极限仍是正态随机变量。

设 $\{W(t),-\infty<t<\infty\}$是方差为 σ^2 的Brown运动，$f(t)$ 和 $g(t)$ 在 $(-\infty,\infty)$ 上连续可微，则由(5.18) 可以得到

$$E\left[\int_a^b f'(t)(W(t)-W(a))\,\mathrm{d}t\int_a^b g'(t)(W(t)-W(a))\,\mathrm{d}t\right]$$
$$=\sigma^2\int_a^b (f(t)-f(b))(g(t)-g(b))\,\mathrm{d}t. \tag{5.19}$$

作为(5.19)式的一个简单应用，令 $a=0$，$b=1$，$f(t)=g(t)=t$，$0\leqslant t\leqslant 1$，则由(5.19)得

$$E\left(\int_0^1 W(t)\,\mathrm{d}t\right)^2=\sigma^2\int_0^1 (t-1)^2\,\mathrm{d}t=\sigma^2/3.$$

由于Brown运动是Gauss过程，故$\int_0^1 W(t)\,\mathrm{d}t$是均值为零，方差为$\sigma^2/3$的正态随机变量。

5.3.2 微分

设二阶矩过程 X 满足上节条件 (i)—(iii)，则过程是均方连续的，即当 $t\to t_0$ 时 $E(X(t)-X(t_0))^2\to 0$，这可由 $E(X(t)-X(t_0))^2=r(t,t)+r(t_0,t_0)-2r(t,t_0)+(m(t)-m(t_0))^2\to 0$ 得出。如果我们进一步假定 X 满足对每个 $\omega\in\Omega$，

(iv) 在 T 的任一有限闭区间上只有有限个第一类间断点，

(v) 存在 $\{Y(t),t\in T\}$，使

$$X(t)-X(t_0)=\int_{t_0}^{t}Y(s)\mathrm{d}s,\ [t_0,t]\subseteq T.$$

则我们称随机过程是可微的，且把过程 $\{Y(t),\ t\in T\}$ 称为随机过程 X 的导数，记为 $\{X'(t),t\in T\}$。故

$$X(t)-X(t_0)=\int_{t_0}^{t}X'(s)\mathrm{d}s. \tag{5.20}$$

由(5.20)知过程 X 的样本轨道是连续的，除了有限个 $X'(t)$ 的不连续点外，均有

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t}=X'(t).$$

在(5.20)两边取数学期望即得

$$m_X(t)-m_X(t_0)=\int_{t_0}^{t}m_X'(s)\mathrm{d}s.$$

因此

$$m_{X'}(t)=\frac{\mathrm{d}m_X(t)}{\mathrm{d}t},\ t\in T.$$

此外，如果 $X(t),\ t\in T$ 是可微的二阶矩过程，则随机变量 $(X(t+h)-X(t))/h$ 均方收敛于 $X'(t)$，即

$$\lim_{h\to 0}E\left|\frac{X(t+h)-X(t)}{h}-X'(t)\right|^2=0,\ t\in T. \tag{5.21}$$

一个有趣的事实是如果 $\{X(t),t\in T\}$ 是可微的宽平稳过程，则 $X(t)$ 与 $X'(t)$ 是不相关的(见第四章习题6)。

值得注意的是Brown运动W是不可微的。在5.1中已从概率上指出了这一点，这儿再从均方收敛的角度来看，由Brown运动定义容易算得

$$E\left(\frac{W(t+h)-W(t)}{h}\right)^2=\frac{\sigma^2}{|h|},$$

其中 σ^2 为方差。从而当 $h\to 0$ 时

$$\lim_{h\to 0}E\left(\frac{W(t+h)-W(t)}{h}\right)^2=\infty, \tag{5.22}$$

若 $W(t)$ 可微，记为 $W'(t)$，$t\in R$，由(5.21)应有

$$\lim_{h\to 0} E\left|\frac{W(t+h)-W(t)}{h}-W'(t)\right|^2=0,\quad t\in R,$$

即

$$\lim_{h\to 0} E\left(\frac{W(t+h)-W(t)}{h}\right)^2=E(W'(t))^2,\qquad t\in R,$$

与(5.22)矛盾。

5.3.3 关于Brown运动的积分

设 $\{W(t),\ -\infty<t<\infty\}$ 是参数为 σ^2 的Brown运动。a 和 b 为两个有限数以及 $f(t)$ 是$[a,b]$上的连续可微函数。由于Brown运动不可微，故在常义下无法定义积分

$$\int_a^b f(t)W'(t)\mathrm{d}t,$$

但如果把上式看作$\int_a^b f(t)\mathrm{d}W(t)$ 并定义此积分为和式

$$\sum_{i=1}^n f(t_i)(W(t_i)-W(t_{i-1}))$$

的极限，其中 $a=t_0<t_1<\cdots<t_n=b$，且

$$\max_{1\leqslant i\leqslant n}(t_i-t_{i-1})\to 0(n\to\infty),$$

则可以在某种意义下定出上述积分值。

下面我们来推导和式极限的显式。为此，注意到 $f(t)$是$[a,b]$上连续可微函数，故

$$\begin{aligned}
&\sum_{i=1}^n f(t_i)(W(t_i)-W(t_{i-1}))\\
&\quad=f(t_n)W(t_n)-f(t_1)W(t_0)-\sum_{i=1}^{n-1}(f(t_{i+1})-f(t_i))W(t_i)\\
&\quad=f(b)W(b)-f(t_1)W(a)-\sum_{i=1}^{n-1}f'(\xi_i)W(t_i)\Delta t_i,
\end{aligned}$$

其中 $\xi_i\in(t_i,t_{i+1})$，$\Delta t_i=t_{i+1}-t_i$，则当 $\lambda\triangleq\max\limits_{1\leqslant i\leqslant n}\Delta t_i\to 0$ 时$\sum\limits_{i=1}^{n-1}f'(\xi_i)$ $\cdot W(t_i)\Delta t_i$ 趋于积分$\int_a^b f'(t)W(t)\mathrm{d}t$，而由 f 的连续性，$f(t_1)\to$

$f(a)$，从而

$$\lim_{\lambda\to0}\sum_{i=1}^{n}f(t_i)(W(t_i)-W(t_{i-1}))$$

$$=f(b)W(b)-f(a)W(a)-\int_a^b f'(t)W(t)\mathrm{d}t,$$

故我们可以定义函数 f 关于Brown运动 W 的积分为

$$\int_a^b f(t)W'(t)\mathrm{d}t=\int_a^b f(t)\mathrm{d}W(t)$$

$$=f(b)W(b)-f(a)W(a)-\int_a^b f'(t)W(t)\mathrm{d}t. \quad (5.23)$$

注意到上式右边是有明确定义的且与通常的分部积分相符。我们称(5.23)式左边为连续可微函数关于Brown运动的随机积分。

由(5.23)，若记 $\mathrm{d}W(t)=W'(t)\mathrm{d}t$，则 $W'(t)$ 就称为Brown运动的"导数"，也称为"白噪声"。由于Brown运动有方差参数σ^2，故对应的白噪声称为有参数 σ^2 的白噪声。但我们要注意它不是常义下的随机过程。它是作为一种泛函指出(5.23)左边积分的值。特别取 $f(t)=1$，我们有

$$\int_{t_0}^{t}W'(s)\mathrm{d}s=\int_{t_0}^{t}\mathrm{d}W(s)=W(t)-W(t_0). \quad (5.24)$$

白噪声 $W'(t)$ 是非常有用的，因为它能描述许多物理系统。在下一小节我们将用它来定义随机微分方程。但是，我们应该牢记白噪声不是真正的导数，它是一类泛函，只有在关于Brown运动的积分中才有意义。另一方面，由于$W'(t)\mathrm{d}t$表示Brown运动的增量 $\mathrm{d}W(t)$，故若以 h 表示 t 的增量，则 $\mathrm{d}W(t)=W(t)-W(t-h)$ 是均值为零，方差为 $h\sigma^2$ 的正态随机变量。由此知

$$\left\{\frac{W(kh)-W((k-1)h)}{h},\quad k\geqslant1\right\}$$

是第四章中定义的Gauss白噪声序列。容易算得，其协方差函数为

$$R(k)=\begin{cases}\sigma^2/h, & k=0,\\ 0, & k>0.\end{cases}$$

其对应的谱密度为 $S(\omega)=\sigma^2/h$。显然，当 $h\to0$ 时协方差函数趋

于 δ-函数。

由于Brown运动是Gauss过程，故由(5.23)知$\int_a^b f(t)W'(t)\mathrm{d}t$是正态随机变量，均值为零。为计算其方差，我们将证明若 $a\leqslant b$，g 是$[a,b]$上另一个连续可微函数，则

$$E\left[\int_a^b f(t)\mathrm{d}W(t)\int_a^b g(t)\mathrm{d}W(t)\right]=\sigma^2\int_a^b f(t)g(t)\mathrm{d}t. \tag{5.25}$$

在上式中令 $g=f$，则

$$\mathrm{Var}\left(\int_a^b f(t)\mathrm{d}W(t)\right)=\sigma^2\int_a^b f^2(t)\mathrm{d}t. \tag{5.26}$$

特别取 $f=1$，则 $\mathrm{Var}\left(\int_a^b \mathrm{d}W(t)\right)=\sigma^2(b-a)$，与 $W(b)-W(a)$ 的方差相同。

现在我们证明(5.25)式。由(5.23)，

$$\int_a^b f(t)\mathrm{d}W(t)=f(b)(W(b)-W(a))-\int_a^b f'(t)(W(t)-W(a))\mathrm{d}t$$

把 f 换成 g，可得到 g 的类似等式，因此

$$\begin{aligned}&E\left[\int_a^b f(t)\mathrm{d}W(t)\int_a^b g(t)\mathrm{d}W(t)\right]\\&=E\left[f(b)(W(b)-W(a))-\int_a^b f'(t)(W(t)-W(a))\mathrm{d}t\right]\\&\quad\cdot\left[g(b)(W(b)-W(a))-\int_a^b g'(t)(W(t)-W(a))\mathrm{d}t\right].\end{aligned}$$

把期望号内项乘开并分成四项，交换期望号和积分号，并注意到

$$\begin{aligned}&E\left[(W(b)-W(a))\int_a^b g'(t)(W(t)-W(a))\mathrm{d}t\right]\\&=\int_a^b g'(t)\sigma^2(t-a)\mathrm{d}t=\sigma^2\left((b-a)g(b)-\int_a^b g(t)\mathrm{d}t\right),\end{aligned}$$

结合习题 8，我们有

$$E\left[\int_a^b f(t)\mathrm{d}W(t)\int_a^b g(t)\mathrm{d}W(t)\right]$$
$$= f(b)g(b)(b-a)\sigma^2 - f(b)g(b)(b-a)\sigma^2$$
$$+ f(b)\sigma^2\int_a^b g(t)\mathrm{d}t - g(b)f(b)(b-a)\sigma^2$$
$$+ g(b)\sigma^2\int_a^b f(t)\mathrm{d}t + \sigma^2\int_a^b (f(b)-f(t))(g(b)-g(t))\mathrm{d}t$$
$$= \sigma^2\int_a^b f(t)g(t)\mathrm{d}t.$$

利用Brown运动的独立增量性，关于Brown运动的随机积分还有如下两个有用的结果，即

$$E\left[\int_a^b f(t)\mathrm{d}W(t)\int_c^d g(t)\mathrm{d}W(t)\right] = 0, \quad a \leqslant b \leqslant c \leqslant d, \quad (5.27)$$

$$E\left[\int_a^b f(t)\mathrm{d}W(t)\int_a^c g(t)\mathrm{d}W(t)\right]$$
$$= \sigma^2\int_a^b f(t)g(t)\mathrm{d}t, \ a \leqslant b \leqslant c, \quad (5.28)$$

这儿 f 和 g 是上述积分区间上的两个连续可微函数。证明留给读者。

例 5.1 设 $\{X(t), t \geqslant 0\}$ 定义为

$$X(t) = \int_0^t e^{\alpha(t-u)}\mathrm{d}W(u), \quad 0 \leqslant t < \infty.$$

其中 α 为实常数。求过程 X 的均值和协方差函数。

解 由(5.23)知 $EX(t) = 0$。由(5.28)，对 $0 \leqslant s \leqslant t$，

$$EX(t)X(s) = E\int_0^s e^{\alpha(s-u)}\mathrm{d}W(u)\int_0^t e^{\alpha(t-u)}\mathrm{d}W(u)$$
$$= e^{\alpha(s+t)} E\left(\int_0^s e^{-\alpha u}\mathrm{d}W(u)\int_0^t e^{-\alpha u}\mathrm{d}W(u)\right)$$
$$= \sigma^2 e^{\alpha(s+t)}\int_0^s e^{-2\alpha u}\mathrm{d}u$$
$$= \sigma^2(e^{\alpha(s+t)} - e^{\alpha(t-s)})/2\alpha.$$

因此由对称性得

$$r_X(s,t)=\sigma^2(e^{\alpha(s+t)}-e^{\alpha|t-s|})/2\alpha,\quad s\geqslant 0,\ t\geqslant 0.$$

特别，令 $s=t$ 得

$$\mathrm{Var}(X(t))=\sigma^2(e^{2\alpha t}-1)/2\alpha,\quad t\geqslant 0.$$

在(5.23)中，如果 f 是 $(-\infty,b]$ 上的连续可微函数，且

$$\int_{-\infty}^{b} f^2(t)\mathrm{d}t<\infty,$$

定义

$$\int_{-\infty}^{b} f(t)\mathrm{d}W(t)=\lim_{a\to-\infty}\int_{a}^{b} f(t)\mathrm{d}W(t),$$

则 $\int_{-\infty}^{b} f(t)\mathrm{d}W(t)$ 以概率 1 有限，且是均值为零，方差为 $\sigma^2\int_{-\infty}^{b} f^2(t)\mathrm{d}t$ 的正态分布。若 $g(t)$ 是 $(-\infty,c]$ 上连续可微函数，且 $\int_{-\infty}^{c} g^2(t)\mathrm{d}t<\infty$，则

$$E\int_{-\infty}^{b} f(t)\mathrm{d}W(t)\int_{-\infty}^{c} g(t)\mathrm{d}W(t)$$
$$=\sigma^2\int_{-\infty}^{\min(b,c)} f(t)g(t)\mathrm{d}t.$$

进一步，在 $\int_{-\infty}^{\infty} f_i^2(t)\mathrm{d}t<\infty$，$i=1,2$ 的情况下，定义

$$\int_{-\infty}^{\infty} f(t)\mathrm{d}W(t)=\lim_{\substack{a\to-\infty\\ b\to+\infty}}\int_{a}^{b} f_i(t)\mathrm{d}W(t),\quad i=1,2,$$

则仍可证明 $\int_{-\infty}^{\infty} f(t)\mathrm{d}W(t)$ 以概率 1 有限，且是均值为零，方差为 $\sigma^2\int_{-\infty}^{\infty} f_i^2(t)\mathrm{d}t$ 的正态随机变量，$i=1,2$，而且成立

$$E\int_{-\infty}^{\infty} f_1(t)\mathrm{d}W(t)\int_{-\infty}^{\infty} f_2(t)\mathrm{d}W(t)=\sigma^2\int_{-\infty}^{\infty} f_1(t)f_2(t)\mathrm{d}t.$$

5.3.4 随机微分方程

1．随机微分方程的引入

我们在本章开始时引入 Brown 运动这一模型来描述粒子由

于分子碰撞而产生的运动情况。如果用随机微分方程，我们可以对这类物理过程用另外的模型来描述。

设 $X(t)$ 表示在时刻 t，一沿直线运动的质点的位置。(或者可设 $X(t)$ 表示质点空间位置的一个坐标，则 $X'(t)$ 和 $X''(t)$ 分别表示在时刻 t 时质点的速度和加速度。再以 m 表示质点的质量以及 $F(t)$ 表示在时刻 t 作用在质点上的力。由牛顿第二运动定律

$$F(t)=mX''(t). \tag{5.29}$$

考虑三种类型的外力

(1) 由介质粘性产生的摩擦力 $-fX'(t)$，它正比于速度而方向与速度方向相反；

(2) 弹性力 $-kX(t)$，如在摆或弹簧中受到的力，它与到平衡位置的距离成正比，方向指向平衡位置；

(3) 外力 $\xi(t)$，它与质点运动无关。

为方便起见，我们考虑如下形式的力

$$F(t)=-fX'(t)-kX(t)+\xi(t), \tag{5.30}$$

这儿 f 和 k 为非负常数，由(5.29)和(5.30)得微分方程

$$mX''(t)+fX'(t)+kX(t)=\xi(t). \tag{5.31}$$

假定外力 ξ 由某个随机因素引起，则我们可以把 $\{\xi(t),t\in T\}$ 作为一个随机过程。此时 $X(t),t\in T$ 也是一个随机过程而(5.31)是与这两个过程有关的微分方程，称为随机微分方程。如果外力来自于分子的碰撞，则由物理背景知这种外力可看作是参数为 σ^2 的白噪声。此时 $X(t)$ 是满足如下随机微分方程的随机过程

$$mX''(t)+fX'(t)+kX(t)=W'(t). \tag{5.32}$$

由于白噪声不是通常的随机过程，所以(5.32)也不是通常意义下的微分方程，其含义将在下面小节中讨论。



图 5.3

除了与分子碰撞直接有关的领域外，在其他领域中也有大量的含白噪声的微分方程。例如考虑图 5.3 的简单电路。它由电阻 R，电感 L 和电容 C

组成。以 $X(t)$ 表示在时刻 t 电容 C 两端的电压降，由克希霍夫第二定律知 $X(t)$满足微分方程：

$$LCX''(t)+RCX'(t)+X(t)=\xi(t).$$

即使在没有电源的情况下，由于电阻中电子的热运动仍在电路中产生一个小电压源，称为"热噪声"，由物理学知这个热噪声也是白噪声的形式，此时电压降满足随机微分方程

$$LCX''(t)+RCX'(t)+X(t)=W'(t).$$

当然，考虑更复杂的电子系统，我们也可得到更高阶的随机微分方程，例如

$$a_0X^{(n)}(t)+a_1X^{(n-1)}(t)+\cdots+a_nX(t)=W'(t),$$

这儿 $a_0,\cdots,a_n$ 为实常数且 $a_0\neq 0$，而 $W'(t)$ 是参数为 σ^2 的白噪声。

2．一阶随机微分方程

考虑一阶常系数线性随机微分方程

$$a_0X'(t)+a_1X(t)=W'(t), \tag{5.33}$$

其中 a_0 和 a_1 为实常数，且 $a_0\neq 0$，$W'(t)$ 是参数为 σ^2 的白噪声。如上一小节所言(5.33)实际上是没有定义的，因为白噪声不能作为通常意义下的有样本函数的随机过程。但由于 $\mathrm{d}W(t)=W'(t)\mathrm{d}t$，所以(5.33)可以写为

$$a_0X'(t)\mathrm{d}t+a_1X(t)\mathrm{d}t=\mathrm{d}W(t),$$

两边从 t_0 到 t 积分得

$$a_0(X(t)-X(t_0))+a_1\int_{t_0}^{t}X(s)\mathrm{d}s=W(t)-W(t_0). \tag{5.34}$$

方程 (5.34) 称为随机积分方程，它是有定义的。因为对每个 $\omega\in\Omega$，Brown运动的样本函数 $W(t)=W(t,\omega)$ 是有定义的连续函数，现设 $X(t)$ 定义在区间 I 上，$t_0\in I$，且 $X(t_0)$ 是已知的随机变量，则我们称满足 (5.34) 的有连续样本函数的随机过程 X 是随机微分方程(5.33)在区间 I 上的解。

由以上定义知随机微分方程(5.33)的实际意义是随机积分方

程(5.34)。为了与通常的常微分方程相对应，也为了方便起见，我们常用(5.33)的微分形式。

在通常的一阶非齐次线性微分方程求解中，常用常数变易法。这种方法稍作变动后可以用来求(5.33)的通解。以下记$\alpha = -a_1/a_0$，对应于(5.33)的齐次通解为

$$X(t) = Ce^{\alpha t},$$

现用常数变易法，令$C = C(t)$，则

$$X'(t) = (C'(t) + \alpha C(t))e^{\alpha t},$$

代入(5.33)得

$$C'(t)e^{\alpha t} = \frac{W'(t)}{a_0}, \tag{5.35}$$

即

$$dC(t) = \frac{1}{a_0}e^{-\alpha t}dW(t).$$

两边从t_0到t积分，由(5.23)得

$$\begin{aligned} C(t) - C(t_0) &= \frac{1}{a_0}\int_{t_0}^{t} e^{-\alpha s}dW(s) \\ &= \frac{1}{a_0}(e^{-\alpha t}W(t) - e^{-\alpha t_0}W(t_0)) \\ &\quad + \int_{t_0}^{t} \alpha\, e^{-\alpha s}W(s)ds. \end{aligned}$$

由$X(t_0) = C(t_0)e^{\alpha t_0}$得$C(t_0) = e^{-\alpha t_0}X(t_0)$，代入上式经过简单整理得

$$\begin{aligned} X(t) = \left(X(t_0) - \frac{W(t_0)}{a_0}\right)e^{\alpha(t-t_0)} + \frac{W(t)}{a_0} \\ + \frac{\alpha}{a_0}\int_{t_0}^{t} e^{-\alpha(s-t)}W(s)ds, \quad t \in I. \end{aligned} \tag{5.36}$$

即随机积分方程(从而(5.33))的解有(5.36)右边的形式。反之，由于上述推导均为可逆，故(5.36)右边的过程满足(5.34)。因此(5.36)式表示了(5.33)的通解。易见(5.36)右边可改写为

$$X(t)=X(t_0)e^{\alpha(t-t_0)}+\frac{1}{a_0}\int_{t_0}^{t}e^{\alpha(t-s)}dW(s),\quad t\in I. \tag{5.37}$$

这与通常微分方程解的结构相符，即(5.33)的通解等于对应的齐次方程的通解加上(5.33)的一个特解。不同之处在于这儿的 $X(t_0)$ 是随机变量而不一定是常数。此外，(5.37)右边是关于Brown运动的随机积分，它由(5.33)定义。在许多应用问题中，初始值 $X(t_0)$ 恰是一个常数 x_0，这时(5.37)可写为

$$X(t)=x_0e^{\alpha(t-t_0)}+\frac{1}{a_0}\int_{t_0}^{t}e^{\alpha(t-s)}dW(s),\quad t\in I.$$

这是一个 Gauss 过程。下面计算其均值函数和协方差函数。为简单计，设 $t_0=0$，$X(t)$ 是$[0,\infty)$上满足初值 $X(0)=x_0$ 的随机微分方程(5.33)的解，即

$$X(t)=x_0e^{\alpha t}+\frac{1}{a_0}\int_0^t e^{\alpha(t-s)}dW(s),\quad t\geqslant 0.$$

由于对白噪声的积分的均值为零，故

$$EX(t)=x_0e^{\alpha t},\quad t\geqslant 0.$$

由(5.28)并注意到 $-2\alpha a_0^2=2a_0a_1$，易得

$$r_X(s,t)=\frac{\sigma^2}{2a_0a_1}\left(e^{\alpha|t-s|}-e^{\alpha(t+s)}\right),\quad s\geqslant 0,\ t\geqslant 0.$$

特别

$$\mathrm{Var}(X(t))=\frac{\sigma^2}{2a_0a_1}\left(1-e^{2\alpha t}\right),\quad t\geqslant 0.$$

下面讨论 $\alpha=-a_1/a_0<0$ 的情况，定义

$$X_0(t)=\frac{1}{a_0}\int_{-\infty}^{t}e^{\alpha(t-s)}dW(s),\quad -\infty<t<\infty.$$

则当 $-\infty<t<\infty$ 时

$$\begin{aligned}X_0(t)&=\frac{1}{a_0}\int_{-\infty}^{0}e^{\alpha(t-s)}dW(s)+\frac{1}{a_0}\int_0^t e^{\alpha(t-s)}dW(s)\\&=X_0(0)e^{\alpha t}+\frac{1}{a_0}\int_0^t e^{\alpha(t-s)}dW(s),\end{aligned}\tag{5.38}$$

这就是(5.37)中 $t_0=0$ 的形式。因此 $\{X_0(t),-\infty<t<\infty\}$ 在

$(-\infty,\infty)$上满足(5.33)。可以证明 $X_0(t)$ 是在$(-\infty,\infty)$上满足随机微分方程(5.33)的唯一的宽平稳过程。如果在$[0,\infty)$上来考虑满足(5.33)的二阶矩过程，则可以证明该过程渐近等于$(-\infty,\infty)$上满足(5.33)的那个宽平稳解 $X_0(t)$。

由常数变易法，我们也可以进一步考虑如下变系数一阶线性随机微分方程

$$a_0X'(t)+a_1(t)X(t)=W'(t),\quad t\in I, \tag{5.39}$$

其中 $a_0(t)$ 和 $a_1(t)$ 分别是区间 I 上的连续可微函数和连续函数且在 I 上 $a_0(t)\neq 0$，我们可得到(5.39)的通解为

$$X(t)=\left(X(t_0)-\frac{W(t_0)}{a_0(t_0)}\right)\exp\left\{-\int_{t_0}^{t}p(u)\mathrm{d}u\right\}+\frac{W(t)}{a_0(t)}$$
$$-\int_{t_0}^{t}\exp\left\{\int_{s}^{t}p(u)\mathrm{d}u\right\}\frac{a_1(s)-a_0'(s)}{a_0^2(s)}W(s)\mathrm{d}s,$$

其中 $p(t)=a_1(t)/a_0(t)$。

5.3.5 n阶随机微分方程

在上一小节我们讨论了一阶线性随机微分方程，这一小节我们继续讨论 n 阶线性随机微分方程。

$$a_0(t)X^{(n)}(t)+a_1(t)X^{(n-1)}(t)+\cdots+a_n(t)X(t)$$
$$=W'(t),\quad t\in I, \tag{5.40}$$

这儿 $a_0(t),\cdots,a_n(t)$ 是区间 I 上的连续函数，且在 I 上$a_0(t)\neq 0$，$W'(t)$是参数为 σ^2 的白噪声。如同一阶线性随机微分方程(5.33)一样的理由，(5.40)是没有定义的，我们说随机过程 X 是在含有 t_0 的区间 I 上微分方程(5.40)的一个解是指过程 X 是 $n-1$次可微的，且在 I 上满足

$$\int_{t_0}^{t}(a_0(t)X^{(n)}(t)+a_1(t)X^{(n-1)}(t)+\cdots+a_n(t)X(t))\mathrm{d}t$$
$$=W(t)-W(t_0). \tag{5.41}$$

由非随机线性常微分方程的一般理论知道，若(5.40)右边是非随机连续函数，则解为 $X^*(t)+Y(t)$ 这种形式，其中 $X^*(t)$

是(5.40)的一个特解，而 $Y(t)$ 为对应齐次方程的通解，它可表为 n 个线性无关的齐次解的线性组合。解的这种结构也可以平移到随机微分方程(5.40)。即若 $X_1(t),\cdots,X_n(t)$ 是

$$a_0(t)X^{(n)}(t)+\cdots+a_n(t)X(t)=0,\quad t\in I \tag{5.42}$$

的 n 个线性无关解，$C_1,\cdots,C_n$ 是任意 n 个随机变量，则(5.42)的解 $Y(t)$ 可表为

$$Y(t)=C_1X_1(t)+\cdots+C_nX_n(t),\quad t\in I. \tag{5.43}$$

若随机过程 $\{X^*(t),t\in I\}$ 满足(5.40)(即在(5.41)意义下)，则满足(5.40)的随机过程的一般形式为

$$X(t)=C_1X_1(t)+\cdots+C_nX_n(t)+X^*(t). \tag{5.44}$$

由于(5.42)同非随机线性微分方程相同，所以其 n 个线性无关解就是通常非随机意义下的解。因此关键问题是求出(5.40)的一个特解。由于一般情形下求解比较复杂，故下面仅讨论系数 $a_0(t),\cdots,a_n(t)$ 为常数 $a_0,\cdots,a_n$ 时(5.40)的特解。我们用脉冲响应法。首先简要回忆一下非随机线性微分方程解的结构。设齐次常系数线性微分方程为

$$a_0X^{(n)}(t)+\cdots+a_nX(t)=0, \tag{5.45}$$

其中 $a_0\neq 0$。我们称 $a_0k^n+a_1k^{n-1}+\cdots+a_n=0$ 为(5.45)的特征方程。则一定存在满足(5.45)的 n 个线性无关解 $\phi_1(t),\cdots,\phi_n(t)$，满足

$$\phi_j^{(k)}(0)=\begin{cases}1, & k=j-1,\\ 0, & 0\leqslant k\leqslant n-1\text{ 且 }k\neq j-1.\end{cases} \tag{5.46}$$

若给定初值 $X(t_0),X'(t_0),\cdots,X^{(n-1)}(t_0)$，则(5.45)的解可表为

$$X(t)=X(t_0)\phi_1(t-t_0)+\cdots+X^{(n-1)}(t_0)\phi_n(t-t_0), \tag{5.47}$$

定义

$$h(t)=\begin{cases}0, & t<0,\\ \phi_n(t)/a_n, & t\geqslant 0.\end{cases} \tag{5.48}$$

显然 $h(0)=h'(0)=\cdots=h^{(n-2)}(0)=0$，$h^{(n-1)}(0)=1/a_n$。若 $f(t)$ 是连续函数，令

$$X(t)=\int_{t_0}^{t}h(t-s)f(s)\mathrm{d}s,$$

则 $X(t)$ 在 $[t_0,\infty)$ 上满足非齐次线性微分方程

$$a_0X^{(n)}(t)+\cdots+a_nX(t)=f(t),\quad t\geqslant t_0,\tag{5.49}$$

以及初始条件 $X(t_0)=\cdots=X^{(n-1)}(t_0)=0$ 的解。如果(5.49)右边是稳定的(即特征方程的根都在左半平面内)，则当 $t\to\infty$时$h(t)$以指数速度趋于零。如果连续函数 $f(t)$，当 $t\to-\infty$时增长不太快，例如

$$\int_{-\infty}^{\infty}\mathrm{e}^{ct}|f(t)|\mathrm{d}t<\infty\quad 对所有\ c>0,$$

则

$$X(t)=\int_{-\infty}^{t}h(t-s)f(s)\mathrm{d}s=\int_{-\infty}^{\infty}h(t-s)f(s)\mathrm{d}s$$

是满足(5.49)的在 $(-\infty,\infty)$ 上的解。当 $f(t)$ 是0时刻的单位脉冲时，(5.49)的解

$$X(t)=\int_{-\infty}^{\infty}h(t-s)f(s)\mathrm{d}s=h(t),$$

即 $h(t)$ 是在0时刻输入的单位脉冲在 t 时刻的响应。

现在回到求常系数随机线性微分方程

$$a_0X^{(n)}(t)+a_1X^{(n-1)}(t)+\cdots+a_nX(t)=W'(t)\tag{5.50}$$

的特解上来。我们有如下的重要结论：

定理 5.2 设 $h(t)$ 由(5.48)定义，则随机过程

$$X(t)=\int_{t_0}^{t}h(t-s)\mathrm{d}W(s),\quad t\geqslant t_0\tag{5.51}$$

是(5.50)在$[t_0,\infty)$上满足初始条件

$$X(t_0)=\cdots=X^{(n-1)}(t_0)=0\tag{5.52}$$

的一个解。

证明可参见文献[3]。

由(5.44)和(5.47)以及定理 5.2，我们得知在$[t_0,\infty)$上满足(5.49)及初始条件 $X(t_0)=\cdots=X^{(n-1)}(t_0)=0$ 的一般解为

$$X(t)=X(t_0)\phi_1(t-t_0)+\cdots+X^{(n-1)}(t_0)\phi_n(t-t_0)$$

$$+\int_{t_0}^{t} h(t-s)\mathrm{d}W(s), \quad t\geqslant t_0. \tag{5.53}$$

如果 $X(t_0), \cdots, X^{(n-1)}(t_0)$ 是 n 个非随机的实常数，记为 $c_1, \cdots, c_n$，则(5.53)右边是正态随机变量，其均值和方差函数为

$$EX(t)=c_1\phi_1(t)+\cdots c_n\phi_n(t), \tag{5.54}$$

$$\operatorname{Var}(X(t))=\sigma^2\int_0^t h^2(t-s)\mathrm{d}s=\sigma^2\int_0^t h^2(s)\mathrm{d}s. \tag{5.55}$$

如果(5.50)左边是稳定的(即对应的特征方程的根在左半平面内)，则随机过程

$$X_0(t)=\int_{-\infty}^{t} h(t-s)\mathrm{d}W(s), \quad -\infty<t<\infty \tag{5.56}$$

有定义，且在$(-\infty,\infty)$上满足(5.50)式。且随机过程 X_0 是均值为 0 的 Gauss 过程，其协方差函数为

$$r_{X_0}(s,t)=\sigma^2\int_{-\infty}^{\min(s,t)} h(s-u)h(t-u)\mathrm{d}u$$

$$=\sigma^2\int_{-\infty}^{\infty} h(s-u)h(t-u)\mathrm{d}u,$$

因此 X_0 是宽平稳的，其协方差函数

$$R_{X_0}(t)=\sigma^2\int_{-\infty}^{\infty} h(-u)h(t-u)\mathrm{d}u, \quad -\infty<t<\infty$$

可以证明此随机过程 X_0 是在 $(-\infty, \infty)$ 上满足 (5.50) 的唯一宽平稳过程。

此外，以宽平稳过程 Y 代替白噪声，我们也可以考虑如下的随机线性微分方程

$$a_0X^{(n)}+a_1X^{(n-1)}(t)+\cdots+a_nX(t)=Y(t), \tag{5.57}$$

则同样可以证明，随机过程

$$\int_{t_0}^{t} h(t-s)Y(s)\mathrm{d}s, \quad t\geqslant t_0$$

是满足(5.57)的一个解，从而(5.57)的一般解为

$$X(t)=C_1\phi_1(t)+\cdots+C_n\phi_n(t)+\int_{t_0}^{t} h(t-s)Y(s)\mathrm{d}s,$$

其中 $\phi_1(t), \cdots, \phi_n(t)$ 由(5.46)定义。如果用 $Y(t)+W'(t)$ 代替

(5.57)的右端，则求其一个特解相应于分别求(5.50)和(5.57)的特解 $X_1^*(t)$ 和 $X_2^*(t)$，其和 $X_1^*(t)+X_2^*(t)$ 即为所求的特解。

例 5.2 考虑二阶线性随机微分方程

$$X''(t)+2X'(t)+2X(t)=W'(t) \tag{5.58}$$

(1) 设 $X(t)$，$0\leqslant t<\infty$ 是该随机微分方程在 $[0,\infty)$ 上的解，满足初始条件 $X(0)=0$ 和 $X'(0)=1$。求在使 $EX(t)=0$ 的最小正时刻 t 时 $X(t)$ 的分布。

(2) 求(5.58)在 $(-\infty,\infty)$ 上的宽平稳解，并求出使 $X_0(0)$ 和 $X_0(t)$ 不相关的最小正时刻 t。

解 (5.58)对应的特征方程为

$$k^2+2k+2=0,$$

故两根为 $k=-1\pm i$，由 $X(0)=0$ 和 $X'(0)=1$ 可求出满足(5.46)的 $\phi_1(t)$ 和 $\phi_2(t)$ 为

$$\phi_1(t)=e^{-t}(\cos t+\sin t),\quad \phi_2(t)=e^{-t}\sin t$$

由(5.48)得

$$h(t)=\begin{cases}0, & t<0,\\ e^{-t}\sin t, & t\geqslant 0.\end{cases}$$

由(5.53)知满足给定初始条件的随机过程为

$$X(t)=\int_0^t e^{-(t-s)}\sin(t-s)\,dW(s),\quad t\geqslant 0, \tag{5.59}$$

这是一个Gauss过程。由(5.54)和(5.55)得

$$EX(t)=\phi_2(t)=e^{-t}\sin t,$$

$$\begin{aligned}\mathrm{Var}(X(t))&=\sigma^2\int_0^t h^2(s)\,ds=\sigma^2\int_0^t e^{-2t}\sin^2 t\,dt\\ &=\frac{\sigma^2}{8}[1+e^{-2t}(\cos 2t-\sin 2t-2)].\end{aligned}$$

由上知使 $EX(t)=0$ 的最小正时刻是 $t=\pi$。而 $X(\pi)$ 服从均值为0，方差为 $\sigma^2(1-e^{-2\pi})/8$ 的正态分布。由于 $k=-1\pm i$ 在左半平面内，故由(5.56)知(5.58)的宽平稳解为

$$X_0(t)=\int_{-\infty}^t h(t-s)\,dW(s)=\int_{-\infty}^t e^{-(t-s)}\sin(t-s)\,dW(s).$$

其协方差函数为

$$R_{X_0}(t)=\sigma^2 e^{-|t|}(\cos|t|+\sin|t|)/8.$$

因此使 $X_0(t)$ 和 $X_0(0)$ 不相关的正时刻为 $t=3\pi/4$。

5.4 Brown运动的其他一些应用

上节给出了 Brown 运动在随机微分方程中的一些应用，本节将给出Brown运动在其他一些领域中的应用。

首先，如果 $\{X_n, n\geqslant 1\}$ 是独立同分布随机变量序列，且 $EX_1=0$，$EX_1^2=\sigma^2>0$，令 $S_n=X_1+\cdots+X_n$。以 $[a]$ 记实数 a 的整数部分，则对 $\forall t\in[0,1]$，可以定义随机过程序列 W_n：

$$W_n(t)=S_{[nt]}/\sigma\sqrt{n},\quad 0\leqslant t\leqslant 1.$$

容易看出，对固定的 t，由中心极限定理，当 $n\to\infty$ 时 $W_n(t)$ 的分布趋于均值为 0，方差为 t 的正态分布。故由Brown运动的定义可以推知

定理 5.3 (Donsker)在上面记号下，若记 W 为标准Brown运动，则*

$$W_n \xrightarrow{d} W \quad (n\to\infty)$$

例 5.3 (反正弦律)设随机信号序列 $\{X(n), n=0,1,\cdots\}$ 为独立同分布随机变量序列，且

$$P(X(n)=1)=P(X(n)=-1)=1/2,\quad n=1,2,\cdots$$

令

$$S(n)=\sum_{i=1}^{n}X(i),$$

即到时刻 n 为止正信号领先负信号的次数。我们来求在时刻 j 到时刻 k 这段时间内，正信号领先权不改变的概率。即若 $S(j)>0$，

* 这儿 $\{W_n(t), 0\leqslant t\leqslant 1\}$ 趋于标准 Brown 运动 $W=\{W(t), 0\leqslant t\leqslant 1\}$ 是指对 $W_n(t)$ 的任意有限维分布趋于对应标准Brown运动的有限维分布

则 $S(j+1),\cdots,S(k)$ 都非负;若 $S(j)<0$,则 $S(j+1),\cdots,S(k)$ 皆非正的概率。即 $S(i)$, $j<i<k+1$, 符号相同的概率。精确计算很麻烦,这儿我们用 Brown 运动来近似计算。由定理 5.2, $W_n(t)=2S_{[nt]}/\sqrt{n}$ 趋于标准 Brown 运动,故上述概率 $P(S(i),$ $j<i<k+1$,符号相同)与 $P(W_n(t)$ 不变号, $j/n<t<(k+1)/n)$ 相同,而这一概率可以用 $P(W(t)$ 不变号, $j/n<t<(k+1)/n)$ 来近似。即在时间区间 $(j/n,(k+1)/n)$ 中 Brown 运动不过 0 点的概率。一般,若以 $A(t_1,t_2)$ 表示在 (t_1,t_2) 中 Brown 运动至少过一次 0 点的概率。由于 $W(t_1)$ 服从 $N(0,t_1)$,故由条件期望的公式得

$$P\{A(t_1,t_2)\}=E[P(A(t_1,t_2)|W(t_1)=x)]$$
$$=\frac{1}{\sqrt{2\pi t_1}}\int_{-\infty}^{\infty}P\{A(t_1,t_2)|W(t_1)=x\}e^{-x^2/2t_1}dx. \tag{5.60}$$

当给定 $W(t_1)=x$ 时, $W_1(t)=W(t+t_1)-W(t_1)$, $t\geqslant 0$ 仍为由第一节所定义的 Brown 运动,因此在 $W(t_1)=x>0$ 的条件下事件 $A(t_1,t_2)$ 发生就相当于 Brown 运动 W_1 到时刻 t_2-t_1 为止时 W_1 的值至少有一次低于 $-x$ 这一事件发生。由首达时的定义,这一事件又可表示为 $\{T_{|x|}<t_2-t_1\}$, $x<0$ 时得相同表达式。故

$$P\{A(t_1,t_2)\}=\frac{1}{\sqrt{2\pi t_1}}\int_{-\infty}^{\infty}P\{T_{|x|}<t_2-t_1\}e^{-x^2/2t_1}dx$$
$$=\frac{2}{\sqrt{2\pi t_1}}\int_0^{\infty}P\{T_{|x|}<t_2-t_1\}e^{-x^2/2t_1}dx$$
$$=\frac{2}{\pi\sqrt{t_1(t_2-t_1)}}\int_0^{\infty}e^{-x^2/2t_1}dx\int_x^{\infty}e^{-y^2/2(t_2-t_1)}dy, \tag{5.61}$$

上面最后等式由(5.15)得到。

在(5.61)中作积分变换

$$x=\sqrt{2t_1}\rho\cos\varphi,\quad y=\sqrt{2(t_2-t_1)}\,\rho\sin\varphi,\quad 0\leqslant\varphi\leqslant\frac{\pi}{2},$$

再注意到 $tg^{-1}\sqrt{\dfrac{t_1}{t_2-t_1}}=\sin^{-1}\sqrt{\dfrac{t_1}{t_2}}$，则 (5.61) 式可得到如下的显式表达式：

$$P\{A(t_1,t_2)\}=1-\frac{2}{\pi}\sin^{-1}\sqrt{\frac{t_1}{t_2}}. \tag{5.62}$$

取 $t_1=xt_2$，$0<x<1$，则有

$$P\{\text{Brown 运动在}(xt_2,t_2)\text{ 中无零点}\}=\frac{2}{\pi}\sin^{-1}\sqrt{x}. \tag{5.63}$$

例如，当 $x=0.025$，则 P\{Brown 运动在 (0.025，1)中无零点\}≈0.1，如在随机信号中，如果 $n=1000$，则 P\{在(25,1000)中领先权不改变\}≈0.1，即平均来说，从第 25 次开始到最终，领先权不改变这种情况在 10 次中就可能发生一次，这一事实也经常用于两个候选人在计票站中计算甲领先于乙的事件概率的计算。而在著名的赌徒公平赌博中也常用此式来近似计算和解释。例如上面的分析就解释了为什么赌徒有时“运气”特别好，领先权保持不变。公式(5.63)称为反正弦律。

除了(5.63)的反正弦律外，还有另一个反正弦律。设 $W(t)$ 为标准 Brown 运动，以 $A(t)$ 表示在 $[0,t]$ 中 Brown 运动处于 t 轴上方的总时间，则对 $0<x<1$，我们有

$$P\{A(t)/t\leqslant x\}=\frac{2}{\pi}\sin^{-1}\sqrt{x}. \tag{5.64}$$

例如把一个均匀硬币连续掷 365 天，设每秒钟掷一次，则用 (5.64) 式近似可计算得当 $x=0.9969$ 时，P\{使 $S(n)>0$ 的 n 的个数/总的投掷次数 $\leqslant 0.9969$\} $=0.95$，其中 $S(n)$ 为到第 n 次为止正面领先于反面的次数。如果把 $S(n)$ 看到赌徒甲到第 n 局时领先于赌徒乙的次数，则上式表明，如果按每秒钟投掷一次硬币计算，则平均来说，每 20 次中就有一次会出现赌徒甲领先时间超过 363 天 20 小时。这是一个非常令人惊讶的结果。如果不作这样的概率分析，很少有人会相信这一事实。

最后我们简要介绍一下 Brown 桥运动在经验分布函数研究

中的应用。

首先设总体 F 为(0,1)上的均匀分布，以 $X_1, X_2, \cdots, X_n$ 记从总体 F 中抽取的一组随机样本。以 $N_n(s)$ 记落在 $(0,s]$ 中的样本个数，即

$$N_n(s)=\sum_{i=1}^{n}I(X_i\leqslant s)=\sum_{i=1}^{n}I_{(0,s]}(X_i).$$

令

$$F_n(s)=N_n(s)/n,\quad 0\leqslant s\leqslant 1, \tag{5.65}$$

则 $F_n(s)$ 称为经验分布函数。现在研究当 n 趋于无穷时，$F_n(s)$ 的极限性质。由大数法则知

$$F_n(s)\xrightarrow{p}EI_{(0,s]}(X_1)=F(s)=s.$$

由 De Moivre-Laplace 局部极限定理，对固定的 s，$0\leqslant s\leqslant 1$，有

$$\sqrt{n}\,(F_n(s)-s)\xrightarrow{d}N(0,s(1-s)) \tag{5.66}$$

注意，上式中极限与 s 有关，即缺少一致性，因此一个现实的问题就是在 (5.66) 中能否得出收敛性与 s 无关？(这在实际应用中是很重要的)。回答是肯定的。若以 $\alpha_n(s)$ 记 $\sqrt{n}\,(F_n(s)-s)$，则 $\alpha_n=\{\alpha_n(s), 0\leqslant s\leqslant 1\}$ 是一个随机过程。进一步我们可以证明过程 α_n 以分布收敛于 Brown 桥过程 $B=\{B(t), 0\leqslant t\leqslant 1\}$。为此我们先验证 $E\alpha_n(s)\to 0$ 及 $E\alpha_n(s)\alpha_n(t)\to s(1-t)$，$0\leqslant s\leqslant t\leqslant 1$ $(n\to\infty)$。即 α_n 的一阶矩和二阶矩分别收敛于 Brown 桥的一阶矩和二阶矩。但若 $X\sim U(0,1)$，则

$$EI(X\leqslant s)=P(X\leqslant s)=F(s)=s,$$

因此

$$E\alpha_n(s)=E\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}I(X_i\leqslant s)-s\right)=\sqrt{n}\,(s-s)=0.$$

设 $0\leqslant s\leqslant t\leqslant 1$，则

$$E\alpha_n(s)\alpha_n(t)=\frac{1}{n}E\sum_{i,j=1}^{n}(I(X_i\leqslant s)-s)(I(X_j\leqslant t)-t)$$

$$= \frac{1}{n} E\sum_{i=1}^{n}(I(X_i \leqslant s) - s)(I(X_i \leqslant t) - t)$$

$$+ \frac{1}{n} E\sum_{i \neq j}(I(X_i \leqslant s) - s)(I(X_j \leqslant t) - t)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(I(X_i \leqslant s) - sI(X_i \leqslant t) - tI(X_i \leqslant s) + st,$$

$$+ \frac{1}{n}\sum_{i \neq j} E(I(X_i \leqslant s) - s) \cdot E(I(X_j \leqslant t) - t)$$

$$= s - st - st + st + 0 = s(1-t).$$

关于过程 α_n 以分布收敛于 Brown 桥的严格数学证明，读者可以参阅有关随机过程收敛性的书或文献，这儿不再细述了。

如果总体分布不是(0,1)上的均匀分布，我们也可以考虑类似的问题。设总体分布 F 是连续函数，随机变量 X 的分布函数为 F，则 $F(X)$ 为(0,1)上的均匀分布。这是由于 F 为连续增函数，故当 $0 \leqslant y \leqslant 1$ 时，

$$P(F(X) \leqslant y) = P(X \leqslant F^{-1}(y)) = F(F^{-1}(y)) = y.$$

当 $y<0$ 或 $y>1$ 时，$P(F(X) \leqslant y)$ 的值分别为 0 和 1，由均匀分布的定义即得。

现设总体分布 F 连续，$X_1, X_2, \cdots X_n$ 为从总体 F 中抽取的一组样本量为 n 的一组样本，由上面分析知，$F(X_1), F(X_2), \cdots, F(X_n)$ 为 (0, 1) 上均匀分布的一组随机样本。对 $F(X_1), \cdots, F(X_n)$ 进行类似讨论，令 $\alpha_n(s) = \sqrt{n}(F_n(s) - F(s))$，则 $\alpha_n = \{\alpha_n(s), 0 \leqslant s \leqslant 1\}$ 以分布收敛于 Brown 桥 B。

由Brown桥的定义，对每个 $t \in (0,1)$，$B(t)$的分布服从均值为0，方差为 $t(1-t)$ 的正态分布，进一步还可以算出 $\sup_{0 \leqslant t \leqslant 1} B(t)$，$\sup_{0 \leqslant t \leqslant 1} |B(t)|$ 等 Brown 桥泛函的分布函数。例如

$$P(\sup_{0 \leqslant t \leqslant 1} B(t) \leqslant y) = \begin{cases} 1 - e^{-2y^2}, & y \geqslant 0, \\ 0, & y < 0. \end{cases} \tag{5.67}$$

$$P(\sup_{0\le t\le 1}|B(t)|\le y)=\begin{cases}1-2\sum_{k=1}^{\infty}(-1)^{k+1}e^{-2k^2y^2}, & y\geq 0,\\ 0, & y<0.\end{cases} \tag{5.68}$$

由于 $\{\alpha_n(s),\ 0\leqslant s\leqslant 1\}$ 以分布收敛于 Brown 桥，因此作为 $\alpha_n(s)$ 的连续线性泛函 $\sup\limits_{0\le t\le 1}\alpha_n(s)$ 及 $\sup\limits_{0\le t\le 1}|\alpha_n(s)|$ 也以分布收敛于 $\sup\limits_{0\le t\le 1}B(t)$ 以及 $\sup\limits_{0\le t\le 1}|B(t)|$。利用这一事实，可以得出著名的 Glivenlko 定理：

$$\sup_{0\le t\le 1}|F_n(s)-F(s)|\xrightarrow{p}0\quad (n\to\infty), \tag{5.69}$$

其中分布函数 F 为连续函数。此外，在统计学的假设检验中，利用(5.67)和(5.68)可以检验一个未知的分布是否是某个连续分布这一问题。例如，给了你一组样本，但不知道这组样本的总体是否就是正态总体，这时利用 $\sup\limits_{0\le t\le 1}\alpha_n(t)$ 以及 $\sup\limits_{0\le t\le 1}|\alpha_n(t)|$ 的收敛性以及(5.67)和(5.68)式，就可以检设这个假设。这种检验分别称为 Smirnov 和 Kolmogorov 检验。它们在数理统计中是两个极为重要的检验。

有关 Brown 运动及其应用的更进一步的叙述，读者可以参见文献[1][2][3]。

习 题 五

1. 设 $\{W(t),\ t\geqslant 0\}$ 为标准 Brown 运动。

(1) 求 $W(1)+W(2)+\cdots+W(n)$ 的分布，

(2) 证明Brown运动的增量过程 $\{Y(t)=W(t+1)-W(t),\ t\geqslant 0\}$ 是平稳过程，并证明其谱密度函数为 $\dfrac{1-\cos\omega}{\pi\omega^2}$。

2. 设 $\{W(t),\ t\geqslant 0\}$ 为 Brown 运动，验证下列过程仍为 $[0,\infty)$ 上的 Brown 运动。

(1) $Y(t)=tW(1/t)$,

(2) $W(t)=W(a^2t)/a$, $a>0$。

3. 设 $\{W(t), t\geqslant 0\}$ 为 Brown 运动，$W(t)\sim N(0,\sigma^2 t)$，令 $Y(t)=W(t)+\mu t$，则称 $\{Y(t), t\geqslant 0\}$ 为带漂移的 Brown 运动。试确定 $Y(t)$ 的分布，并求 $EY(t)$ 和 $\mathrm{Var}(Y(t))$。

4. {设 $W(t), t\geqslant 0\}$ 为标准Brown运动，令 $Z(t)=|W(t)|$，$t\geqslant 0$。求 $Z(t)$ 的分布以及 $EZ(t)$ 和 $\mathrm{Var}(Z(t))$。

5. 设$\{Z(t), t\geqslant 0\}$为Brown桥。验证$\left\{W(t)=(t+1)Z\left(\frac{t}{t+1}\right), t\geqslant 0\right\}$是 Brown 运动。

6. 验证$\left\{B(t)=(1-t)W\left(\frac{t}{1-t}\right), 0\leqslant t\leqslant 1\right\}$是Brown桥，其中 $\{W(t), 0\leqslant t\leqslant 1\}$ 是标准 Brown 运动。

7. 设 $\{W(t), t\geqslant 0\}$ 是 Brown 运动，a 为正常数。证明 $\{W(t)-W(a), t\geqslant a\}$ 是 Brown 运动。

8. 设 $\{W(t), -\infty<t<\infty\}$ 是方差为 σ^2 的 Brown 运动，$f(t)$ 和 $g(t)$ 是有限区间 $[a,b]$ 上的两个连续可微函数。证明

$$E\int_a^b f'(t)(W(t)-W(a))\mathrm{d}t\int_a^b g'(t)(W(t)-W(a))\mathrm{d}t$$
$$=\sigma^2\int_a^b (f(b)-f(t))(g(b)-g(t))\mathrm{d}t.$$

提示：交换期望号和积分号并注意到

$E(W(t)-W(a))(W(s)-W(a))=(\min(t,s)-a)\sigma^2$.

以下 $\{W(t), -\infty<t<\infty\}$ 均表示方差为 σ^2 的 Brown 运动。

9. 求 $W(t)$ 和$\int_0^1 W(s)\mathrm{d}s$ 的相关系数，其中 $0\leqslant t\leqslant 1$。

10. 令

$$X(t)=\int_t^{t+1}(W(s)-W(t))\mathrm{d}s,\quad -\infty<t<\infty,$$

证明这是均值为 0 的宽平稳过程并求出 $R_X(t)$。

11．设 $X=\int_0^1 t\mathrm{d}W(t)$，$Y=\int_0^1 t^2\mathrm{d}W(t)$，求出 X 和 Y 的均值和方差，并求 $\mathrm{Cov}(X,Y)$。

12．设 m 和 f 为正常数。$V_0(t)$，$-\infty<t<\infty$ 是随机微分方程

$$mV'(t)+fV(t)=W'(t)$$

的宽平稳解。

(1) 根据白噪声表示 $V_0(t)$，

(2) 求 $V_0(t)$ 的均值和协方差函数，

(3) 令 $X_0(t)=\int_0^t V_0(s)\mathrm{d}s$，$t\geqslant 0$，证明 $\{X_0(t),\ t\geqslant 0\}$ 满足如下随机微分方程

$$mX''(t)+fX'(t)=W'(t),\quad t\geqslant 0,$$

(4) 根据白噪声表示 $X_0(t)$，并求 $X_0(t)$ 的均值和方差。

13．设 c 是实常数。求解如下随机微分方程

$$a_0X'(t)+a_1X(t)=c+W'(t)$$

其中 c，a_0，a_1 为实常数且 $a_0\neq 0$，如果 $a_1/a_0>0$，求上述随机微分方程在 $(-\infty,\infty)$ 上的宽平稳解。

14．在下面的随机微分方程中，求 $\mathrm{Var}(X(t))$，$t\geqslant 0$，其中初始条件均为 $X(0)=X'(0)=0$。如果方程左边是稳定的，求出宽平稳解的协方差函数 $R(t)$，$-\infty<t<\infty$。

(1) $X''(t)+X'(t)=W'(t)$，

(2) $X''(t)+3X'(t)+2X(t)=W'(t)$，

(3) $4X''(t)+8X'(t)+5X(t)=W'(t)$，

(4) $X''(t)+2X'(t)+X(t)=W'(t)$，

(5) $X''(t)+X(t)=W'(t)$。

15．设 $\{X(t),\ -\infty<t<\infty\}$ 满足随机微分方程

$$a_0X'(t)+a_1X(t)=W'(t),$$

而$\{Y(t), -\infty<t<\infty\}$满足如下的随机微分方程：

$$b_0Y'(t)+b_1Y(t)=X(t),$$

证明过程Y满足随机微分方程

$$a_0b_0Y''(t)+(a_0b_1+a_1b_0)Y'(t)+a_1b_1Y(t)=W'(t).$$

16. 设宽平稳过程$\{Y(t), -\infty<t<\infty\}$的均值函数$m_Y=1$以及协方差函数$R_Y(t)=e^{-|t|}$，$-\infty<t<\infty$

(1) 求满足随机微分方程

$$X'(t)+X(t)=Y(t)$$

的宽平稳解$\{X_0(t), -\infty<t<\infty\}$的均值和协方差函数。

(2) 求满足初始条件$X(0)=0$的上述随机微分方程的解$\{X(t), t\geqslant 0\}$的均值和协方差函数。

(3) 验证

$$\lim_{s,t\to+\infty}(r_X(s,t)-r_{X_0}(s,t))=0.$$

17. 设某选区要从候选人甲和乙中选出一位人大代表。民意测验表明，甲乙两人当选的可能性都是50%。设该选区有5000个选举人参加投票。问在计票到2000张时，一个候选人的选票数始终领先于另一个候选人的概率。

附表　常用随机变量的分布与矩母函数

离散概率分布	$P(X=x)$	矩母函数$g(t)$	EX	$\mathrm{Var}X$
二项分布$B(n,p)$, $0\leqslant p\leqslant 1$	$\binom{n}{x}p^x(1-p)^{n-x}$, $x=0,1,\cdots,n$	$(pe^t+(1-p))^n$	np	$np(1-p)$
Poisson分布, $\lambda>0$	$e^{-\lambda}\frac{\lambda^x}{x!}$, $x=0,1,2,\cdots$	$\exp\{\lambda(e^t-1)\}$	λ	λ
几何分布, $0\leqslant p\leqslant 1$	$p(1-p)^{x-1}$, $x=1,2,\cdots$	$\frac{pe^t}{1-(1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
负二项分布, 参数为r,p	$\binom{x-1}{r-1}p^r(1-p)^{x-r}$, $x=r,r+1,\cdots$	$\left(\frac{pe^t}{1-(1-p)e^t}\right)^r$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

连续概率分布	密度函数$f(x)$	矩母函数$g(t)$	EX	$\mathrm{Var}X$
均匀分布$U(a,b)$	$\frac{1}{b-a}$, $a<x<b$	$\frac{e^{ta}-e^{tb}}{t(a-b)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
指数分布, 参数$\lambda>0$	$\lambda e^{-\lambda x}$, $x\geqslant 0$	$\frac{\lambda}{\lambda-t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Γ分布 $\Gamma(n,\lambda)$, $\lambda>0$	$\frac{\lambda e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!}$, $x\geqslant 0$	$\left(\frac{\lambda}{\lambda-t}\right)^n$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$
正态分布, $N(\mu,\sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}$, $-\infty<x<\infty$	$\exp\left\{\mu t+\frac{\sigma^2t^2}{2}\right\}$	μ	σ^2
Beta 分布	$cx^{a-1}(1-x)^{b-1}$, $0<x<1$		$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$
$B(a,b)$, $a>0,b>0$	$c=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$			

参考文献

[1] Sheldom M. Ross, *Stochastic Processes*, 1983, John Wiley & Sons.

[2] Samuel Karlin, Howard M. Taylor, *A First Course in Stochastic Processes*, 1975, Academic Press.

[3] Paul G. Hoel, Sidney C. Port, Charles J. Stone, *Introduction to Stochastic Processes*, 1972, Houghton Mifflin Company.

参 考 文 献

[1] Sheldom M. Ross, *Stochastic Processes*, 1983, John Wiley & Sons.

[2] Samuel Karlin, Howard M. Taylor, *A First Course in Stochastic Processes*, 1975, Academic Press.

[3] Paul G. Hoel, Sidney C. Port, Charles J. Stone, *Introduction to Stochastic Processes*, 1972, Houghton Mifflin Company.